

Align seg 1/1 to: AB006713 from: 1 to: 2699

```

1 MetSerPheGlnGlyLysLysSerIleProArgIleThrSerAspArgLe 17
149 ATGTCCTTCAGAGGCAAGAAAGCATCCCGGGATCAGGATGACCGCT 198
17 uLeuIleArgGlyArgIleValAlaAspGlnSerPheTyrAlaA 34
199 TCTGATCAGAGGTGGAGGATCGTGAATGACGACAGTCTTTACGCTG 248
34 spValHisValGluAspGlyLeuIleLysGlnIleGlyAlaSerLeu 50
249 ATGTGCACGTGGAGATGGCTGTGATAAACAATCGGAGAAACCTCATC 298
51 ValProGlyGlyIleHisThrIleAlaHisGlyLeuMetValLeuPr 67
299 GTCCCTGGGGCATCAGACCATTCAGCCCGACCGCTGTATGGCTCTCC 348
67 oGlyGlyValAspValHisThrArgLeuGlnMetProValLeuGlyMet 84
349 TGGTGGGCTTGTACCTGCACACAGGCTGCAGATGCCTGTCTGGGCATGA 398
84 hrProAlaAspAspPheCysGlnGlyThrLysAlaAlaLeuAlaGlyCly 100
399 CACCGGCTGAGGACTTCTGTGAGGGCACCAGGCGAGCGCTAGCAGGAGGA 448
101 ThrThrMetIleLeuAspHisValPheProAspThrGlyValSerLeuLe 117
449 ACCACCATGATCTTGGACCAAGCTTCTCCCGACACAGGCTGTGAGCTTGC 498
117 uAlaAlaTyrGluGlnTrpArgGluArgAlaAspSerAlaAlaCysCys 134
499 GCGGGCTACAGCGGTGGCGGGAGCGGGGACAGCGCGCTGTGTGCG 548
134 spTyrSerLeuHisValAspIleThrArgTrpHisGluSerIleLysGlu 150
549 ACTACTCCCTGCACGTGGACATCACCAGATGGCATGAGAGCATCAAGGAG 598
151 GluLeuGluAlaLeuValLysGlyLysValAlaSerPheLeuValPh 167
599 GAGCTGGAGGCCCTGTGTAAGGAGAAGGGTGTGAACCTCTCTCTGCTCT 648
167 eMetAlaTyrLysAspArgCysGlnCysSerAspSerGlnMetTyrGluI 184
649 CATGGCATACAGGACCGGTGCCAGTGCAGCGACAGCCAGATGTACGAGA 698
184 lePheSerIleLeuArgAspLeuGlyAlaLeuAlaGlnValHisAlaGlu 200
699 TCTTCAGCATCATCCGGGACCTGGGGCCCTTGGCCAGCTGCACGCTGAG 748
201 AsnGlyAspIleValGluGluGlnLysArgLeuLeuGluGlyIle 217
749 AACGGGACATCGTGAGGAGGAGCAGAGCGGTGCTGGAGCTCGGCAT 798
217 eThrGlyProGluGlyHisValSerHisProGluGluValGluAlaG 234
799 CACTGGCCCGAGGCGCAGGTCTCAGCCACCCCGAGGAGGTGGAGGCTG 848
234 luAlaValTyrArgAlaValThrIleAlaLysGlnAlaAsnCysProLeu 250
849 AGCGGGTGTACCGAGCTGTTCACCATCGCCAGCAGCAACCTGCCCGCTG 898
251 TyrValThrLysValMetSerLysGlyAlaAlaAspAlaIleAlaGlnAl 267
899 TACGTCACCAAGGTGATGAGCAGGGGGCGCGGACCATCGCTCGCTCAGC 948
267 aLysArgArgGlyValValPheGlyGluProIleThrAlaSerLeuG 284
949 CAAGCGCAGAGGGTGTGCTGTTTGGGAGCCCATCACCAGCCGCTGG 998
284 lyThrAspGlySerHisTyrTrpSerLysAsnTrpAlaLysAlaAla 300

```

```

999 GCACCGACGGTTTCACACTACTGGAGCAAGAACTGGGCCAAGGCCGACGCC 1048
301 PheValThrSerProProValAsnProAspProThrThrAlaAspHisLe 317
1049 TTCGTCACATCACCCCTGTCAACCCACAGCCACCGCAGCAGACCACT 1098
317 uThrCysLeuLeuSerSerGlyAspLeuGlnValThrGlySerAlaHisC 334
1099 CACCTGTCTGTGTCCAGCGGGGACCTCCAGGTGACAGCAGCGCCCACT 1148
334 ystThrPheThrThrAlaGlnLysAlaValGlyLysAspAsnPheAlaLeu 350
1149 GCACCTTCACCACTGCCAGAGGCTGTGGCAAGGACCACTTCGCGCTG 1198
351 IleProGluGlyThrAsnGlyIleGluGluArgMetSerMetValTrpG1 367
1199 ATCCCGAGGGCAGCAACCGCATTTGAGGAGCGCATGTCGATGGTCTGGA 1248
367 uLysCysValAlaSerGlyLysMetAspGluAsnGluPheValAlaValT 384
1249 GAAATGTGTGGCCCTCTGGGAAGATGGAGCAGAGATGAGTTCGCGGTGA 1298
384 hrSerThrAsnAlaAlaLysIlePheAsnPheTyrProArgLysGlyArg 400
1299 CCAGTACAAATGCTGCCAAATCTTCAATTTTACCAAGGAAGGGCGGA 1348
401 ValAlaValGlySerAspAlaAspLeuValIleTrpAsnProLysAlaTh 417
1349 GTGGCTGTGGCTCTGACGCTGACCTGTGCATATGGAACCCCAAGGCCAC 1398
417 rLysIleLeuSerAlaLysThrHisAsnLeuAsnValGluTyrAsnIleP 434
1399 CAGATCATCTCTGCCAAGACCCCAATCTGAACGTGGAGTACAACTCT 1448
434 heGluGlyValGluCysArgGlyAlaProAlaValIleSerGlnGly 450
1449 TCGAGGAGGTGGAGTCCCGGGAGCGCTCCCGTGGTGCATAGTCAGGCC 1498
451 ArgValAlaLeuGluAspGlyLysMetPheValThrProGlyAlaGlyAr 467
1499 CGAGTGGCGCTGGAGGACGGGAAGATGTTTGTACCCCGGGGCGGCCG 1548
467 gPheValProArgLysThrPheProAspPheValTyrLysArgIleLysA 484
1549 CTTCTGCTCCCTCGGAAACATTCCTCCGACCTTGTCTACAGAGGATCAA 1598
484 laArgAsnArgLeuAlaGluIleHisGlyValProArgGlyLeuTyrAsp 500
1599 CTCGCAACAGGCTGGCGGAGATCCACGCTGTGCCCGTGGCTGTATGAC 1648
501 GlyProValHisGluValMetValProAlaLysProGlySerGlyAlaPr 517
1649 GGGCCGCTCCAGAGGTGATGGTCCCTGCCAAGCCAGGAGGTGGCGCTCC 1698
517 alaArgAlaSerCysProGlyLysIleSerValProProValArgAsnL 534
1699 GGCCGCGCGCTCTGCCAGGCAAGATCTCCGTGCTCTCTCTCTCTCTCT 1748
534 euHisGlnSerGlyPheSerLeuSerGlySerGlnAlaAspHisIle 550
1749 TACATCAGTCGGGGTTCAGCCCTATCTGGGTCTCAGGCTGATGACCACATC 1798
551 AlaArgArgThrAlaGlnLysIleMetAlaProProGlyGlyArgSerAs 567
1799 GCCCGACGCACAGCAGAGATCATGGCACCTGGCGGCGCGCTCCAA 1848
567 nIleThrSerLeuSer 572
1849 CATCACCTCTCTCTCC 1864
seq_name: gb_pat1:A87376
seq_documentation_block:

```

LOCUS A87376 1690 bp DNA PAT 22-JAN-2000
DEFINITION Sequence 7 from Patent WO9837192.
ACCESSION A87376
VERSION A87376.1 GI:6736129
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 1690)
AUTHORS Byk,T. and Belin,M.
TITLE USE OF ULIP PROTEINS IN THE DIAGNOSIS AND THERAPY OF CANCER AND
PARANEOPLASTIC NEUROLOGICAL SYNDROMES
JOURNAL Patent: WO 9837192-A 7 27-AUG-1998;
BYK TAMARA (CH); BELIN MARIE FRANCOISE (FR)
FEATURES
Location/Qualifiers
1..1690
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 360 a 505 c 539 g 286 t
ORIGIN

alignment_scores:
Quality: 2882.00 Length: 553
Ratio: 5.221 Gaps: 0
Percent Similarity: 99.819 Percent Identity: 99.819

alignment_block:
US-09-367-496-8 x A87376 ..
Align seg 1/1 to: A87376 from: 1 to: 1690

1 MetSerPheGlnGlyLysLysSerIleProArgIleThrSerAspArgLe 17
|||||
32 ATGCTCTCCAGGCAAGAAAGATCCCGCGGATCAGAGTGACCGCT 81
|||||
17 uLeuIleArgGlyGlyArgIleValAsnAspAspGlnSerPheTyrAla 34
|||||
82 TCGATCAGAGGTGGGAGATCGTGAATGACGACAGTCTCTTTAGCTG 131
|||||
34 spValHisValGluAspGlyLeuIleLysGlnIleGlyGluAsnLeu 50
|||||
132 ATGTCAGCGTGGAGATGGCTTGTATAAAACAAATCGAGAAACCTCATC 181
|||||
51 ValProGlyLylleHisThrIleAspAlaHisGlyLeuMetValLeuPr 67
|||||
182 GTCCCTGGGGGCATCAGACCATGACGCCACGGCTGATGCTCTTCC 231
|||||
67 oGlyGlyValAspValHisThrArgLeuGlnMetProValLeuGlyMet 84
|||||
232 TGGTGGGCTTGAGCTCCACACAGGCTGCAGATGCCCTGTCTGGCATGA 281
|||||
84 hrProAlaAspAspPheCysGlnGlyThrLysAlaAlaLeuAlaGlyGly 100
|||||
282 CACCGGCTGACGACTTGTCTGAGGCGACCAAGGCGCTAGCAGAGGA 331
|||||
101 ThrThrMetIleLeuAspHisValPheProAspThrGlyValSerLeu 117
|||||
332 ACCACCATGATCTGGACACAGCTCTCCCGACACGGGTGTGAGCTGCT 381
|||||
117 uAlaAlaTyrGluGlnTrpArgGluArgAlaAspSerAlaAlaCysCys 134
|||||
382 GCGCGCTACGAGCAGTGGCGGAGCGGCGGACAGCGCGCTGCTGCG 431
|||||
134 spTyrSerLeuHisValAspIleThrArgTrpHisGluSerIleLysGlu 150
|||||
432 ACTACTCCCTGCAGTGGACATCAACCGATGGCATGAGATCAAGGAG 481
|||||
151 GluLeuGluAlaLeuValLysGluLysGlyValAsnSerPheLeuValPh 167
|||||
482 GAGCTGGAGGCGCTGGTCAAGGAGAGGGTGTGAACCTCTCTGCTGCTT 531
|||||

167 eMetAlaTyrLysAspArgCysGlnCysSerAspSerGlnMetTyrGluI 184
|||||
532 CATGGCATACAAAGGACCGGTGCCAGTGCAGCGACGACGATGTACGAGA 581
|||||
184 lePheSerIleIleArgAspLeuGlyAlaLeuAlaGlnValHisAlaGlu 200
|||||
582 TCTTCAGCATCATCCGGACCTGGGGGCTTGGCCAGGTGCACGCTGAG 631
|||||
201 AsnGlyAspIleValGluGluGluLysArgLeuLeuGluLeuGlyI 217
|||||
632 AACGGGACATCTGGAGGAGGAGCAGAGCGGTGCTGGAGCTCGGCAT 681
|||||
217 eThrGlyProGluGlyHisValLeuSerHisProGluGluValGluAlaG 234
|||||
682 CAC"GGCCCGGAGGCGCACGTGCTCAGCCACCCCGAGAGGTGGAGCTG 731
|||||
234 luAlaValTyrArgAlaValThrIleAlaLysGlnAlaAsnCysProLeu 250
|||||
732 AGGGGTGTACCGAGCTGTCCACCATGCCAAGCAGGCAAACTGCCCGCTG 781
|||||
251 TyrValThrLysValMetSerLysGlyAlaAlaAspAlaIleAlaGlnAl 267
|||||
782 TACGTACCAAGGTGTAGAGCAAGGGGCGCGCCAGCCATCGCTCAGGC 831
|||||
267 aLysArgArgGlyValValPheGlyGluProIleThrAlaSerLeuG 284
|||||
832 CAAGCGCAGAGGGTGGTGGTGGTGGGAGCCCATCACCGCAGCCTGG 881
|||||
284 lyThrAspGlySerHisTyrTrpSerLysAsnTrpAlaLysAlaAla 300
|||||
882 GCACCGACGGTTCACACTACTGGAGCAAGAACTGGGCCAAGGCTGCAGCC 931
|||||
301 PheValThrSerProValAsnProAspProThrThrAlaAspHisLe 317
|||||
932 TTCCTCATCATCACCCCTGTCAACCCACAGCCACCCAGCGACACCT 981
|||||
317 uThrCysLeuLeuSerSerGlyAspLeuGlnValThrGlySerAlaHisC 334
|||||
982 CACCTGTCTGTCTCCAGCGGGACCTCCAGGTGACAGGCGAGCCCACT 1031
|||||
334 ysThrPheThrThrAlaGlnLysAlaValGlyLysAspAsnPheAlaLeu 350
|||||
1032 GCACCTTCACCACTGCCAAGAGGCTGTGGGCAAGGACAACTTCGCGCTG 1081
|||||
351 IleProGluGlyThrAsnGlyIleGluArgMetSerMetValTrpG 367
|||||
1082 ATCCCGGAGGCGACCAACGGCATTCAGGAGCGCATGTCGATGCTCTGGA 1131
|||||
367 uLysCysValAlaSerGlyLysMetAspGluAsnGluPheValAlaValT 384
|||||
1132 GAATGTGTGGCTCTGGGAAGATGGACGAGATGAGTTCTGTCGCGGTGA 1181
|||||
384 hrSerThrAsnAlaAlaLysIlePheAsnPheTyrProArgLysGlyArg 400
|||||
1182 CCAGTACAAATGCTGCCAAATCTCAATTTTACCAAGGAAGGGGCGCA 1231
|||||
401 ValAlaValGlySerAspAlaAspLeuValIleTrpAsnProLysAlaTh 417
|||||
1232 GTGGCTGTGGCTCTGACGCTGACCTGGTGCATATGGAAACCCCAAGGCCAC 1281
|||||
417 rLysIleIleSerAlaLysThrHisAsnLeuAsnValGluTyrAsnIleP 434
|||||
1282 CAAGATCATCTCTGCCAAGACCCCACTGCGCTGCTCATAGTCAACATCT 1331
|||||
434 heGluGlyValGluCysArgGlyAlaProAlaValIleSerGlnGly 450
|||||
1332 TCCAGGAGGTGGAGTCCCGGGAGCGCTGCGCTGCTCATAGTCAAGGCG 1381
|||||
451 ArgValAlaLeuGluAspGlyLysMetPheValThrProGlyAlaGlyArg 467
|||||
1382 CGAGTGGCGCTGGAGGACGGGAAGATGTTGTCCACCCCGGGGCGGGCCG 1431
|||||
467 gPheValProArgLysThrPheProAspPheValTyrLysArgIleLysA 484
|||||

```

|||||
1432 CTTGTCCTCGGAACATTCCCGGACTTGTCTACAAGAGGATCAAG 1481
|||||
484 laArgAsnArgLeuAlaGluIleHisGlyValProArgGlyLeuTyrAsp 500
|||||
1482 CTCGCAACAGGCTGGCGAGATCCACGGTGTGCCCGTGGCTGTATGAC 1531
|||||
501 GlyProValHisGluValMetValProAlaLysProGlySerGlyAlaPr 517
|||||
1532 GGGCCCGTCCACGAGGTGATGGTGCCTGCCAAGCCAGGAGTGGCGTCC 1581
|||||
517 oAlaArgAlaSerCysProGlyLysIleSerValProValArgAsnL 534
|||||
1582 GGCCCGCGCTCGTCCGACGCAAGATCTCCGTGCCTCTGTGCGCAAC 1631
|||||
534 euHisGlnSerGlyPheSerLeuSerGlySerGlnAlaAspAspHisIle 550
|||||
1632 TACATCAGTCGGGTTCAGCTATCTATGCTGCTCAGGCTGATGACCATC 1681
|||||
551 AlaArgArg 553
|||||
1682 GCCCGAGCG 1690

```

seq_name: gb_p-r9:HSRNAULI4

seq_documentation_block:

```

LOCUS HSRNAULI4 1690 bp mRNA PRI 09-MAY-1997
DEFINITION H.sapiens mRNA for Ulip4 protein.
ACCESSION Y10976
VERSION Y10976.1 GI:2077995
KEYWORDS phosphoprotein.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Byk,T., Ozon,S. and Sobel,A.
The Ulip4s: a family of proteins related to the awonal guidance
associated unc-33 gene product
Unpublished
2 (bases 1 to 1690)
Byk,T.
Direct Submission
Submitted (03-FEB-1997) T. Byk, INSERM U440, 17 rue du Fer a
Moulin, F- 75005 Paris, FRANCE
On May 10, 1997 this sequence version replaced gi:1914866.
Related sequence X87817.

```

FEATURES

```

source
1..1690
/organism="Homo sapiens"
/db_xref="taxon:9606"
/tissue_type="retina"
/clone="ATCC No. 469253"
32..>1690
/notes="mistake in reverse transcription 197..199"
/codon_start=1
/product="cytosolic phosphoprotein"
/protein_id="CAA71872.1"
/db_xref="GI:1946203"
/db_xref="SWISS-PROT:O14531"
/translation="MSFGKKSIPRTSDRLIRGRIVNDQSFYADVHVHEDLIQK
IGNLIVPGGIXITIDAGMLVPGVDVHTRLQPMVLGMPADDFCGTKAALAGGTT
MILDFVFPDTSLLAAYEQWRERADSAACCDYSLHVDITRWHEISKEELALVKEK
VNSFLVPMAYKDRCCSDSQMYEFSIIRDLGLAQVHAENGDIIVEEOKRRLLELGI
GPBGHVLSPHEVEAEAVRYVTIAQANCLYVTVMKSGRAADAQAQRKRVVVFG
EPITASLGDGSHYKNAKAAAFVTSPPNDPTADHLTCLSSGLQVTSAGH
TFTTAQAKGDNFALIPGTNGIEERMSWKEKCVASGMDENEFVAVTSTNAKIF
NFYPRKRVAVSGDADLVLTWNPATKIIISAKTHNLNVEYNIPEGVRCGAPAVVISOG
RVALEDKMFVTPGAGRFVPRKTFPDEVYKRIKARNRLAEIHGVPRLYDGPVHEVM
PAKPGSGAPARASCPCKISVPPVNRNLHQSGFSLSGSDQADHIARR"
BASE COUNT 359 a 505 c 539 g 286 t 1 others
ORIGIN
|||||

```

alignment_scores:

Quality: 2882.00 Length: 553
Ratio: 5.221 Gaps: 0
Percent Similarity: 99.819 Percent Identity: 99.819

alignment_block:

US-09-367-496-8 x HSRNAULI4 ..

Align seg 1/1 to: HSRNAULI4 from: 1 to: 1690

```

1 MetSerPheGlnGlyLysSerIleProArgIleThrSerAspArgLe 17
|||||
32 ATGTCTTCCAGGCAAGAAAGCATCCCCGGATCACGAGTGACCGCT 81
|||||
17 uLeuIleArgGlyArgIleValAsnAspAspGlnSerPheTyrAla 34
|||||
82 TCTGATCAGAGTGGGAGGATCGTGAATGACGACCACTCTTTACGCTG 131
|||||
34 spValHisValGluAspGlyLeuIleLysGlnIleGlyGluAsnLeu 50
|||||
132 ATGTGACGTGGAGATGGCTTGATAAAACAAATCGGAGAAACCTCATC 181
|||||
51 ValProGlyIleHisThrIleAspAlaHisGlyLeuMetValLeuPr 67
|||||
182 GTCCCTGGGGCATCNAGACCATTTAGCGCCACGGCTGATGGTCTTCC 231
|||||
67 oGlyGlyValAspValHisThrArgLeuGlnMetProValLeuGlyMet 84
|||||
232 TGGTGGCGTTGACGTCCACAAAGGCTGCAGATGCCTGTCTGGGCATGA 281
|||||
84 hrProAlaAspPheCysGlnGlyThrLysAlaAlaLeuAlaGlyGly 100
|||||
282 CACCGGCTGACGACTTCTGTACGGCACCAAGCAGCGCTAGCAGGAGGA 331
|||||
101 ThrThrMetIleLeuAspHisValPheProAspThrGlyValSerLeu 117
|||||
332 ACCACCATGATCTTGGACACGCTTCTCCCGCACACGGGTGTGAGCGTCT 381
|||||
117 uAlaAlaTyrGluThrArgGluAlaAspSerAlaAlaCysCysA 134
|||||
382 GCGGGCTACGACGAGTGGCGGAGCGGGGACAGCGCGGCTGTGCGG 431
|||||
134 spTyrSerLeuHisValAspIleThrArgThrHisGluSerIleLysGlu 150
|||||
432 ACTACTCCTGACGTGGACATCACCCGATGCGATGAGAGCATCAAGAGG 481
|||||
151 GluLeuGluAlaLeuValLysGluLysGlyValAsnSerPheLeuValPh 167
|||||
482 GAGCTGGAGGCGCTGTCAAGAGAGAGGGGTGTGACTCCTCTGCTGCT 531
|||||
167 eMetAlaTyrLysAspArgCysGlnCysSerAspSerGlnMetTyrGlu 184
|||||
532 CATGCATACAGGACCGGTGCCAGTGCAGCCAGCACCCAGATGTACGAGA 581
|||||
184 lePheSerIleIleArgAspLeuGlyAlaLeuAlaGlnValHisAlaGlu 200
|||||
582 TCTTCAGCATCATCCGGGACCTTGGGGGCGCTTGGCCAGGTGCACGCTGAG 631
|||||
201 AsnGlyAspIleValGluGluGlnLysArgLeuLeuGluLeuGlyIle 217
|||||
632 AACGGGACATCGTGGAGGAGGAGCAGAGCGGTGTGAGGCTCGGCAT 681
|||||
217 eThrGlyProGluGlyHisValLeuSerHisProGluGluValGluAlaG 234
|||||
682 CACTGGCCCCGAGGGCCACGTGTCAGCCACCCCGGAGGAGGAGGCTG 731
|||||
234 luAlaValTyrArgAlaValThrIleAlaLysGlnAlaAsnCysProLeu 250
|||||
732 AGGGCGGTATCCGAGCTGTCACTCGCAAGCAGGACCAACTGCCCGCTG 781
|||||
251 TyrValThrLysValMetSerLysGlyAlaAlaAspAlaIleAlaGlnAl 267
|||||

```


Tue Jul 31 13:08:18 2001

```

782 TACGTACCAAGGTGATGACCAAGGGGGCGCCGACGCCATCGCTCAGGC 831
267 alyArgArgGlyValValValPheGlyGluProIleThrAlaSerLeuG 284
832 CAAGCGCAGAGGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 881
284 lyThrAspGlySerHisThrTrpSerLysAsnTrpAlaLysAlaAla 300
882 GCACCGAGGGTTCACACTTACCTGGAGCAGAACTGGCCCAAGCTGCAGCC 931
301 PheValThrSerProProValAsnProAspProThrThrAlaAspHisLe 317
932 TTCGTTCACATCACCCCTGTCAACCCAGACCCACCCAGCGAGACCACT 981
317 uThrCysLeuLeuSerSerGlyAspLeuGlnValThrGlySerAlaHisC 334
982 CACCTGCTGTGTCCAGCGGGGACCTCAGGTGACAGCGAGCGCCCACT 1031
334 ysThrPheThrThrAlaGlnLysAlaValAlGlyLysAspAsnPheAlaLeu 350
1032 GCACCTTCACACTGCCAGAGAGGTGTGGCAAGGACCACTTCGCGCTG 1081
351 ileProGluGlyThrAsnGlyIleGluGluArgMetSerMetValTrpG1 367
1082 ATCCCGAGGGCACCACCAAGCGCATTCAGGAGCGCATGTCGATGGTGGGA 1131
367 uLysCysValAlaSerGlyLysMetAspGluAsnGluPheValAlaValT 384
1132 GAAATGTGGGCTCTGGGAGATGACGAGGATGAGTTCGTCGCGGTGA 1181
384 hrSerThrAsnAlaAlaLysIlePheAsnPhetyrProArgLysGlyArg 400
1182 CCAGTACAAATGTGCAAAATCTTCAATTTTACCCAGGAGGGCGGA 1231
401 ValAlaValGlySerAspAlaAspLeuValIleTrpAsnProLysAlaTh 417
1232 GTGGCTGTGGGCTCTACGCTGACCTGGTCTATGATGACCAAGCGCAC 1281
417 rLysIleSerAlaLysThrHisAsnLeuAsnValGluTyrAsnIleP 434
1282 CAAGATCATCTGCCAAGACCCACCAATCTGAACGTGGAGTACAACTCT 1331
434 heGluGlyValGluCysArgGlyAlaProAlaValIleSerGlnGly 450
1332 TCGAGGAGTGGAGTGGCGGGAGCGCTGCGGTGGTCAATAGTCAGGCG 1381
451 ArgValAlaLeuGluAspGlyLysMetPheValThrProGlyAlaGlyAr 467
1382 CGAGTGGGCTGGAGGAGCGGGAAGATGTTGTACACCCCGGGGGGCGCG 1431
467 qPheValProArgLysThrPheProAspPheValTyrLysArgIleLysA 484
1432 CTCTGCTCCCTCGGAAACATTCCTGGGACTTGTCTACAAGAGGATCAAG 1481
484 laArgAsnArgLeuAlaGluIleHisGlyValProArgGlyLeuTyrAsp 500
1482 CTCGCAACAGGCTGGCGGAGATCCAGGTGCGCCCGTGGCTGATGAC 1531
501 GlyProValHisGluValMetValProAlaLysProGlySerGlyAlaPr 517
1532 GGGCCGCTCCAGAGGTGATGTGCTGCCAAGCCAGGAGTGGCGCTCC 1581
517 alaArgAlaSerCysProGlyLysIleSerValProProValArgAsnL 534
1582 GGCCCGCGGCTCTGCCAGGCAAGATCTCCGTGCTCTCTGTGCGCAACC 1631
534 euHisGlnSerGlyPheSerLeuSerGlySerGlnAlaAspHisIle 550
1632 TACATCAGTCGGGTCTACGCTATCTGGGTCTCAGGCTGATGACCAATC 1681
551 AlaArgArg 553
1682 GCCCGACGC 1690

```

seq_name: gb_rol:AB006715

seq_documentation_block:
LOCUS AB006715 2546 bp mRNA ROD 23-AUG-1997
DEFINITION Mus musculus mRNA for dihydropyrimidinase related protein 4,
complete cds.

ACCESSION AB006715.1 GI:2342489
VERSION dihydropyrimidinase related protein 4.
KEYWORDS Mus musculus fetus whole embryo cDNA to mRNA.
SOURCE Mus musculus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 2546)
AUTHORS Nonaka,M.
TITLE Direct Submission
JOURNAL Submitted (19-AUG-1997) to the DDBJ/EMBL/GenBank databases. Masaru
Nonaka, Nagoya City University Medical School, Department of
Biochemistry; 1 Kawasumi, Mizuho-cho, Mizuho-ku, Nagoya, Aichi 467,
Japan (E-mail:mnonaka@med.nagoya-cu.ac.jp, Tel:052-853-8145,
Fax:052-842-3955)

REFERENCE 2 (sites)
AUTHORS Hamajima,N., Kato,Y., Kowaki,M., Wada,Y., Sasasaki,M. and Nonaka,M.
TITLE Novel members of dihydropyrimidinase related protein family
JOURNAL Unpublished (1997)
REFERENCE 3 (sites)
AUTHORS Hamajima,N., Matsuda,K., Sakata,S., Tamaki,N., Sasaki,M. and
Nonaka,M.
TITLE A novel gene family defined by human dihydropyrimidinase and three
related proteins with differential tissue distribution
JOURNAL Gene 180 (1-2), 157-163 (1996)
MEDLINE 97128821
FEATURES Location/Qualifiers

1..2546
/organism="Mus musculus"
/db_xref="taxon:10090"
/dev_stage="fetus"
/tissue_type="whole embryo"
52..1770
/codon_start=1
/product="dihydropyrimidinase related protein 4"
/protein_id="BAA21888.1"
/db_xref="GI:2342490"
/translation="MSFGKKKIPRITSDRLIKGGKIVNDQSFHADLVVEDGLIKQ
IGENLIVPGGIKTIDAHGLVLPDGVVHTRLQPNYMGMPADPCOGTKAALAGGTT
MILDRVPDAGVSLAAAYEQWRERADSAACDYLHVDIPRWHESTKEELEALVRDKG
VNSPLVEMAYKDRCCQTDGQIYEIFSLRIDLGAQAQVHAENGDIIEEQRKLLQGLIT
GPEGLVLSHPVEAEAVYRAVTAQKQANPLVYTKVMSKGAADMAQAKRRGVVFG
EPITASLGTDGSHYWSKNWAKAAAFVTPPINPDPTTADHLTSLSSGDLQVTSAGHC
TFTTAQKAVGKDNFTLIPDGVNGIEERMSVYWEKCAVKMDENEFVAVTSTNAKIF
NFYPRKGRVAVGSDADLVINPRAIKVFSKSHNLNVEYNIIEGVECRGVPTVVISOG
RVVLEDGLNLTTPGAGREIPKTPDFVYKRIKARNRLAEIHGVRGLYDGPVHEVML
PAKPGSGTQARASCAGSKISVPPVRNLHQSGLSFSLGSQADDDHIARTAKIMAPGGRS
NITSLS"

2546
/note="46 A nucleotides"
BASE COUNT 558 a 719 c 738 g 531 t
ORIGIN

source

CDS

polyA_site

alignment_scores:

Quality: 2818.00 Length: 572
Ratio: 4.996 Gaps: 0
Percent Similarity: 98.601 Percent Identity: 93.007

alignment_block:

US-09-367-496-x AB006715 ..

Align seg 1/1 to: AB006715 from: 1 to: 2546

1 MetSerPheGlnGlyLysSerIleProArgIleThrSerAspArgLe 17

|||||

52 ATGTCCTTCCAAAGCAAGAGCATTTCCCGGATAAACGAGCGCGCT 101
 17 uLeuileArgGlyGlyArgGlyValAlaAsnAspGlnSerPheThrAlaA 34
 102 TCTCATCAAGAGTGGGAAGATTGTGACCATGACCATCTTTTCATGTG 151
 34 spValHisValGluAspGlyLeuileLysGlnIleGlyGluAsnLeuile 50
 152 ATCTGTATGTGAAGACGCTCTGATTAAACAATTTGGAGAAATCTCATC 201
 51 ValProGlyGlyLeuileHisThrIleAspAlaHisGlyLeuMetValLeu 67
 202 GTCCTTGGGGGATCAAAACCATTCATGCTATGCTGATGCTGATGCTGCC 251
 67 GglyGlyValAspValHisThrArgLeuGlnMetProValLeuGlyMetT 84
 252 TGGGGGAGTTGACGTTCACACCGCGCTCAGATGCTGTGATGGCATGA 301
 84 hrProAlaAspPheCysGlnGlyThrLysAlaAlaLeuAlaGlyGly 100
 302 CCCAGCTGATGATTCTGTGAGGCGACCAAGCGGCTCTAGCAGGGGG 351
 101 ThrThrMetLeuLeuAspHisValPheProAspThrGlyValSerLeu 117
 352 ACCACCATGATATTGGACCATGTCTTCCCTGACCGCTGGTGTGACCTGCT 401
 117 uAlaAlaTyrGluGlnTrpArgGluArgAlaAspSerAlaAlaCysCys 134
 402 GGCAGCTTATGAGCAGTGGGGGAGCAGCAGACAGCGCGCTGCTGTG 451
 134 spTyrSerLeuHisValAspIleThrArgTrpHisGluSerIleLysGlu 150
 452 ACTACTCTTACATGTGACATTCCTCGCTGGCAGCAGACAGCGCGCTGCTGTG 451
 151 GluLeuGluAlaLeuValLysGluLysGlyValAsnSerPheLeuValPh 167
 502 GAGCTGAGGCGCTTAGTCAGGCAAGAGTGTGAATCTCTCTCTGCTCT 551
 167 eMetAlaTyrLysAspArgCysGlnCysSerAspSerGlnMetTyrGlu 184
 552 CATGGCATAACAAGACAGGTCAGTGTACTGACGCTCAGATATATGAA 601
 184 lePheSerIleIleArgAspLeuGlyAlaLeuAlaGlnValHisAlaGlu 200
 602 TCTTCAGCTTCATCCCGGACCTGGAGCTGTGCGCCAGGTGCACGACAA 651
 201 AsnGlyAspIleValGluGluGlnLysArgLeuLeuGluLysGly 217
 652 AATGGGACATCTGTGAGGAGCAACAAGCGCTGCTGGAGCAAGGCAT 701
 217 eThrGlyProGluGlyHisValLeuSerHisProGluGluValGluAlaG 234
 702 CACTGTGCTGAGGGCCATGTGCTCAGCCACCCAGAGAGGTAGAGCGG 751
 234 luAlaValTyrArgAlaValThrIleAlaLysGlnAlaAsnCysProLeu 250
 752 AGGCTGTGTACAGACGATCATTGTCACAGCAGGCGCAACTGCCCACTA 801
 251 TyrValThrLysValMetSerLysGlyAlaAlaAspAlaIleAlaGlnAl 267
 802 TAGCTCACCAAGGTATGAGCAAGGTCAGCTGACATGGTGTGCCAAGC 851
 267 aLysArgArgGlyValValPheGlyGluProIleThrAlaSerLeuG 284
 852 CAAGCGCAGGGGGGTGGTCTTTTGGGAAACCTATCACTGCCACGCTGG 901
 284 lyThrAspGlySerHisTyrTrpSerLysAsnTrpAlaLysAlaAla 300
 902 GCACGTGATGGCTCACACTACTGTGAGCAGACACTGGGCAAGGCTGCAGCC 951
 301 PheValThrSerProValAsnProAspProThrThrAlaAspHisLe 317
 952 TTTGTCACTTCAACCCCTATCAACCCGAGCCCTACTACTGCAGACCACT 1001

317 uThrCysLeuLeuSerSerGlyAspLeuGlnValThrGlySerAlaHisC 334
 1002 CACCTCTCTGTCTCCAGTGGGACCTCCAGGTGACAGGAGTGCACCT 1051
 334 ysThrPheThrThrAlaGlnLysAlaValGlyLysAspAsnPheAlaLeu 350
 1052 GCACCTTCACTACTGCCCCAGAGGCTCTTGGCAAGACAACTTCACACTG 1101
 351 IleProGluGlyThrAsnGlyIleGluGluArgMetSerMetValTrpG 367
 1102 ATCCCCGAGGAGTCAACGGTATAGAGAGCGCATGTCTGTGCTGGGA 1151
 367 uLysCysValAlaSerGlyLysMetAspGluAsnGluPheValAlaValT 384
 1152 GAAATGTGTGGCTTCAGGGAATAATGACAGAGATGAGTTCGTGGCGTGA 1201
 384 hrSerThrAsnAlaAlaLysIlePheAsnPheTyrProArgLysGlyArg 400
 1202 CCAGCACAATGCTGCCAAATCTTCAATTTTACCCAGGAAGGGCGCT 1251
 401 ValAlaValGlySerAspAlaAspLeuValIleTrpAsnProLysAlaTh 417
 1252 GTGGCGTGGCTCTGATGCTGACCTGGTCACTCTGGAACCCAGGCGCAC 1301
 417 rLysIleIleSerAlaLysThrHisAsnLeuAsnValGluTyrAsnIleP 434
 1302 GAAAGTCTTCTCTGCCAGAGCATAACCTGAATGTAGATACAACATCT 1351
 434 heGluGlyValGluCysArgGlyAlaProAlaValIleValIleSerGlnGly 450
 1352 TTGAGGAGTGGAGTGCCGAGGAGTGCCCGGCTGCTATAGTCAGGCGC 1401
 451 ArgValAlaLeuGluAspGlyLysMetPheValThrProGlyAlaGlyAr 467
 1402 AGAGTGGTGTGGAGGACGAAACCTGCTGTCTACTCCAGGGGCTGGCGG 1451
 467 gPheValProArgLysThrPheProAspPheValTyrLysArgIleLysA 484
 1452 CTTCATCCCGGAGACGTTCCCGGACTTTGTCTATAAGAGATAAAGG 1501
 484 laArgAsnArgLeuAlaGluIleHisGlyValProArgGlyLeuTyrAsp 500
 1502 CTCGACACAGGCTAGCAGAGATCCACGGTGTGCTCCAGGCTGTACGAC 1551
 501 GlyProValHisGluValMetValProAlaLysProGlySerGlyAlaPr 517
 1552 GGGCTGTGCATGAAGTGTATTACCTGCCAAGCAGGAAGTGGCACACA 1601
 517 oAlaArgAlaSerCysProGlyLysIleSerValProProValArgAsnL 534
 1602 GGCCCGTGCATCTGTTACGGCAAGATCTCAGTCCACCCGCTGGCAACC 1651
 534 euHisGlnSerGlyPheSerLeuSerGlySerGlnAlaAspHisIle 550
 1652 TGACACAGTGGGGTTCAGCTATCTGGCTCTCAGGCTCAGCATCATATT 1701
 551 AlaArgArgThrAlaGlnLysIleMetAlaProProGlyGlyArgSerAs 567
 1702 GCCAGAGGTACGGCTCAGAGATCATGGCACCCCGGAGGAGCGCTCCAA 1751
 567 nIleThrSerLeuSer 572
 1752 CATCAGCTCTCTTTC 1767

seq_name: gb_pat1:A87374

seq_documentation_block:

LOCUS A87374

DEFINITION Sequence 5 from patent WO9837192.

ACCESSION A87374

VERSION A87374.1

GI:6736128

KEYWORDS

PAT

22-JAN-2000

SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 1920)
 AUTHORS Byk,T. and Belin,M.
 TITLE USE OF ULIP PROTEINS IN THE DIAGNOSIS AND THERAPY OF CANCER AND
 PARANEOPLASTIC NEUROLOGICAL SYNDROMES
 JOURNAL Patent: WO 9837192-A 5 27-AUG-1998;
 BYK TAMARA (CH); BELIN MARIE FRANCOISE (FR)
 FEATURES
 Location/Qualifiers
 1..1920
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 BASE COUNT 440 a 524 c 562 g 394 t
 ORIGIN
 alignment_scores:
 Quality: 2803.00 Length: 572
 Ratio: 4.988 Gaps: 0
 Percent Similarity: 98.252 Percent Identity: 92.657
 alignment_block:
 US-09-367-496-8 x A87374 ..
 Align seg 1/1 to: A87374 from: 1 to: 1920
 1 MetSerPheGlnGlyLysSerIleProArgIleThrSerAspArgLe 17
 102 ATGTCCTTCCCAAGGCAAGAAGAGCATCCCCGGATACAGGACCGGCT 151
 17 uLeuIleArgGlyGlyArgIleValAsnAspGlnSerPheTyrAlaA 34
 152 TCTCATCAAGGTGGGAAGATTGTGAACGATGACCACTCTTTCATGCTG 201
 34 spValHisValGluAspGlyLeuIleLysGlnIleGlyGluAsnLeuLe 50
 202 ATCTGTATGGAAGAGCGTCTGATTAACAATAATGGAGAAATCTCATVC 251
 51 ValProGlyGlyIleHisThrIleAspAlaHisGlyLeuMetValLeuPr 67
 252 GTCCCTGGGGGCATCAAAACCATCGATGCTCATGGCTGATGGTGGTGC 301
 67 oGlyGlyValAspValHisThrArgLeuGlnMetProValLeuGlyMet 84
 302 TGGGGGAGTTGACCTTCACACCGCGCTGCAGATGCCCTGTGATGGGCATGA 351
 84 hrProAlaAspAspPheCysGlnGlyThrLysAlaAlaLeuAlaGlyGly 100
 352 CCCAGCTGATGATTCTGTGACGGGCACCAAGCGGCTCTAGCAGCGGG 401
 101 ThrThrMetIleLeuAspHisValPheProAspThrGlyValSerLeuLe 117
 402 ACCACCATGATATGGACCATGTGTTCTCTGACGCTGGTGTGAGCCTGCT 451
 117 uAlaAlaTyrGluGlnTrpArgGluArgAlaAspSerAlaAlaCysCysA 134
 452 GGCAGCCTATGACAGTGGCGGGGAGGAGCAGACAGCGCGGCTGCTGTG 501
 134 spTyrSerLeuHisValAspIleThrArgTrpHisGluSerIleLysGlu 150
 502 ACTACTCTTACATGTGGACATCTCTCGCTGGCAGCAGACCAACAAAGAA 551
 151 GluLeuGluAlaLeuValLysGluLysGlyValAsnSerPheLeuValPh 167
 552 GAGCTGGAGGCCCTAGTCAGGACAAAGGTGTGAACCTCTCTCTGGTCTT 601
 167 eMetAlaTyrLysAspArgCysGlnCysSerAspSerGlnMetTyrGluI 184
 602 CATGGCATACAAAGGACAGGTGCCAGTGTACTGACGGGTGAGATATATGAA 651
 184 lePheSerIleIleArgAspLeuGlyAlaLeuAlaGlnValHisAlaGlu 200

652 TCTTCAGCTCATCCGGGACCTGGGAGCTGTGGCCCGACGTCAGCAAGAA 701
 201 AsnGlyAspIleValGluGluGlnLysArgLeuLeuGluLeuGlyIle 217
 702 AATGGGACATCTGTGAGGAGGAACAAGACCGCTCTGTGAGCAAGGCAT 751
 217 eThrGlyProGluGlyHisValLeuSerHisProGluGluValGluAlaG 234
 752 CACTGTGTCTGAGGGCCATGTGCTCAGCCACCACCAAGAGAGGTAGAGCGC 801
 234 luAlaValTyrArgAlaValThrIleAlaLysGlnAlaAsnCysProLeu 250
 802 AGGCTGTGTACAGCAGCTCACCATTGCCAAGCAGGCCAACTGCCCACTA 851
 251 TyrValThrLysValMetSerLysGlyAlaAlaAspAlaIleAlaGlnAl 267
 852 TAGCTCACCAAGGTGATGAGCAAGGGTGCAGCTGACATGTTGTCGCCNAGC 901
 267 aLysArgArgGlyValValPheGlyGluProIleThrAlaSerLeuG 284
 902 CAACCGCAGGGGGTGTGCTCTTTGGGAACTTATCCTCCTGCCAGCCTGG 951
 284 lyThrAspGlySerHisTyrTrpSerLysAsnTrpAlaLysAlaAlaA 300
 952 GCATGTATGGCTCACACTACTGTGAGCAAGAAGTGGGCCAAGGCTGCAGCC 1001
 301 PheValThrSerProProValAsnProAspProThrThrAlaAspHisLe 317
 1002 TTTGTCTTCACTTCACTTCACTTCACTTCACTTCACTTCACTTCACTTCACT 1051
 317 uThrCysLeuLeuSerSerGlyAspLeuGlnValThrGlySerAlaHisC 334
 1052 CACTCTCTGCTGTCCAGTGGGGACCTCCAGGTGACAGCAGTGCCTCACT 1101
 334 ystThrPheThrThrAlaGlnLysAlaValGlyLysAspAsnPheAlaLeu 350
 1102 GCATCTTCACTTCACTTCACTTCACTTCACTTCACTTCACTTCACTTCACT 1151
 351 ileProGluGlyThrAsnGlyIleGluGluArgMetSerMetValTrpG 367
 1152 ATCCCGCAGAGTAGTCAACGGTATAGAACGCGCATGTCTGTGCTGGGA 1201
 367 uLysCysValAlaSerGlyLysMetAspGluAsnGluPheValAlaValT 384
 1202 GAAATGTGTGGCTTCAGGGAATAATGGACGAGAAATGAGTTGCTGTCGGTGA 1251
 384 hrSerThrAsnAlaAlaLysIlePheAsnPheTyrProArgLysGlyArg 400
 1252 CCAGCACAAATGCTGCCAAAATCTCAATTTTACCAGGAAGGGCGT 1301
 401 ValAlaValGlySerAspAlaAspLeuValIleTrpAsnProLysAlaTh 417
 1302 GTGGCGTGGCTGTGATGCTGACCTGTGTCATCTGGAACCCAGGGGCCAC 1351
 417 rLysIleIleSerAlaLysThrHisAsnLeuAsnValGluTyrAsnIleP 434
 1352 GAAGTGTCTCTGCCAAGAGCCATAACCTGAATGTAGAGTACACATCT 1401
 434 heGluGlyValGluCysArgGlyAlaProAlaValIleSerGlnGly 450
 1402 TTCAAGAGTGGAGTCCCGGAGGAGTCCCGAGGTGCTGTCATAGTCAAGGC 1451
 451 ArgValAlaLeuGluAspGlyLysMetPheValThrProGlyAlaGlyAr 467
 1452 AGAGTGTGTGGAGGACGGAACCTGCTGTGTCACCTCCAGGGGCTGGCCG 1501
 467 gPheValProArgLysThrPheProAspPheValTyrLysArgIleLysA 484
 1502 CTTCATTTCCCGGAAGACGTTCCCGCATTTGTCTATAGAGCATAAAGG 1551
 484 laArgAsnArgLeuAlaGluIleHisGlyValProArgGlyLeuTyrAsp 500

1552 CTCGCAACAGGCTAGCAGAGATCCAGGTGTCCTCGAGGCTGTACGAC 1601
 501 GlyProValHisGluValMetValProAlaLysProGlySerGlyAlaPr 517
 1602 GGGCTGTGCATGAAGTATCTACCTGCCAAGCAGGAGTGGCACA 1651
 517 oAlaArgAlaSerCysProGlyLysIleSerValProProValArgAsnL 534
 1652 GGCCCGTGCATCCTGTTTCAGCAAGATCTCAGTGCACCCCGTGGCAACC 1701
 534 euHisGlnSerGlyPheSerLeuSerGlySerGlnAlaAspAspHisIle 550
 1702 TGCACCAAGTCGGGTCCTCAGCTACTCTCAGGCTGACGATCACT 1751
 551 AlaArgThrAlaGlnLysIleMetAlaProProGlyGlyArgSerAs 567
 1752 GCCAGAGGTACGGCTCAGAAGATCATGGCACCCCGGAGGACCTCAA 1801
 567 nileThrSerLeuSer 572
 1802 CATCACGCTCTCTTCC 1817

seq_name: gb_rol:MMULIP4

seq_documentation block:
 LOCUS MMULIP4 1920 bp mRNA ROD 18-APR-1997
 DEFINITION M. musculus mRNA for Ulip4 protein.
 ACCESSION Y09079
 VERSION Y09079.1 GI:1915916
 KEYWORDS phosphoprotein; Ulip4 gene.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 1920)
 Byk, T.; Ozon, S. and Sobel, A.
 The Ulips: a family of proteins related to the axonal guidance
 associated unc-33 gene product
 Unpublished
 2 (bases 1 to 1920)
 Byk, T.
 Direct Submission
 Submitted (28-OCT-1996) T. Byk, INSERM U440, 17 rue du Fer a
 Moulin, F- 75005 Paris, FRANCE
 Related sequence: X87817.
 Location/Qualifiers
 1. .1920
 /organism="Mus musculus"
 /sub_species="Harlan Sprague Dawley"
 /strain="ICR outbred strain"
 /db_xref="taxon:10090"
 /dev_stage="newborn"
 /tissue_type="brain"
 /clone_lib="lambda Zap II, Stratagene, # 936309"
 102. .1820
 /gene="Ulip4"
 102. .1820
 /gene="Ulip4"
 /codon_start=1
 /evidence="experimental"
 /product="Ulip4 protein"
 /protein_id="CAA70299.1"
 /db_xref="GI:1945273"
 /db_xref="SWISS-PROT:O35098"
 /translation="MSFGKKSIPIRTSDRLIKGKIVNDQSFHADLYVEDGLIKQ
 IGENLIVPGIKTIDAHGLMWLPGGVDVHTRLQMPVMTPADDFCQGTAKALAGTT
 MILDHVPDAGVSLIAAEQWRDADSACCDYSLHVDIPRWHESTKEELALVRDKG
 VNSLFLVWAYKDRCDQGYEIFSLIRDLGAQVHAENGDIYEEQKRLLGEGIT
 GPEGVLVSHPEVEAEAVYRTIAKQANCPLYTVKVMKGAADVMVAOKRRGVVVF
 TPTASLDGDSHYSKWNKAAAFVTSPPINPDPTADHLTSLSSGLDQVTSAGHC
 TTTTAAQKAVGDNFTLIPVWNGIEERMSVVKCVASGKMDENEFVATSNAKIF
 NFYPRKRVAVGSADLVINPRATKVISAKSHNLNVEYNLFEVGEGRGVTVVTSIQG
 RWLEEDGNLLVTPGAGRFIPRKTTPDFYVKRIKARNRLAEIHGVPRLDGPVHEVML

PAKPGSGTQARASCSGKISVPPVRNLHQSGFSLGSQADDDHARRTAQKINWPPGGRS
 NITSLS"
 BASE COUNT 440 a 524 c 562 g 394 t
 ORIGIN

alignment_scores:
 Quality: 2803.00 Length: 572
 Ratio: 4.988 Gaps: 0
 Percent Similarity: 98.252 Percent Identity: 92.657
 alignment_block:
 US-09-367-496-8 x MMULIP4 ..

Align seg 1/1 to: MMULIP4 from: 1 to: 1920

1 MetSerPheGlnGlyLysSerIleProArgIleThrSerAspArgLe 17
 102 ATGTCTTCCAAAGCAAGAGAGCATTTCCCGGATAACAGCGACCGCT 151
 17 uLeuIleArgGlyGlyArgIleValAsnAspGlnSerPheTyrAlaA 34
 152 TCTCATCAAAAGTGGGAAGATTGTGAACGATGACCATCTTTCATGCTG 201
 34 spValHisValGluAspGlyLeuIleLysGlnIleGlyGluAsnLeuIle 50
 202 ATCTGTATGTGGAGACGGTCTGATTAAACAAATTTGGAGAAATCTCATC 251
 51 ValProGlyGlyIleHisThrIleAspAlaHisGlyLeuMetValLeuPr 67
 252 GTCCCTGGGGCATCAAAACCATCGATGCTCATGGCTGATGTGTGCTGCC 301
 67 oGlyGlyValAspValHisThrArgLeuGlnMetProValLeuGlyMet 84
 302 TGGGGAGTGTGAGTTCACACCCGGCTGCAGATGCTGTGATGGCATGA 351
 84 hrProAlaAspPheCysGlnGlyThrLysAlaAlaLeuAlaGlyGly 100
 352 CCCAGCTGATGATTCTCTCAGGCGACCAAGCGGCTCTAGCAGCGGG 401
 101 ThrThrMetIleLeuAspHisValPheProAspThrGlyValSerLeu 117
 402 ACCACCATGATATTGGACCATGTGTTCTCTGACGCTGTGTGAGCCTGCT 451
 117 uAlaAlaTyrGluGlnTrpArgGluArgAlaAspSerAlaAlaCysCysA 134
 452 GGCAGCCTATGAGCAGTGGCGGAGGAGACAGACGCGGCTGTCTGTG 501
 134 spTyrSerLeuHisValAspIleThrArgTrpHisGluSerIleLysGlu 150
 502 ACTACTCTTACATGTGCATCTCTGCTGCGACGAGACGACCAAGAA 551
 151 GluLeuGluAlaLeuValLysGluLysGlyValAsnSerPheLeuValph 167
 552 GAGCTGGAGGCGCTAGTCAGGACAAAGGTGTGAATCCTCTCTGCTGCTT 601
 167 eMetAlaTyrLysAspArgCysGlnCysSerAspSerGlnMetTyrGlu 184
 602 CATGCATACAAAGGACAGTGCCTGCTGCTGCTGCTGCTGCTGCTGCT 651
 184 lePheSerIleAlaArgAspLeuGlyAlaLeuAlaGlnValHisAlaGlu 200
 652 TCTTCAGCCTCATCCGGACCTGGGAGCTGTGGCCAGGTGAGCGACGAA 701
 201 AsnGlyAspIleValGluGluGlnLysArgLeuLeuGluLeuGlyI 217
 702 AATGGGACATCGTGGAGGAGAACACAAAGCCCTGCTGGAGCAAGGCAT 751
 217 eThrGlyProGluGlyHisValLeuSerHisProGluGluValGluAlaG 234
 752 CACTGGTCTCTGAGGGCCATGTGCTCAGCCACCCAGAGAGAGGTAGAGCGG 801
 234 luAlaValTyrArgAlaValThrIleAlaLysGlnAlaAsnCysProLeu 250

```

|||||
802 AGCGTGTGTACAGAGCACTCACCATTGCCAAGCAGGCACTGCCACTA 851
|||||
251 TyrValThrLysValMetSerLysGlyAlaAlaaspAlaGlnAl 267
|||||
852 TACGTCACCAAGGTGATGAGCAAGGGTGACGTGACATGTTGCCAAGC 901
|||||
267 aLysArgGlyValValPheGlyGluProIleThrAlaSerLeuG 284
|||||
902 CAAGCGCAGGGGGTGGTCTCTTGGGGAACCTATCCTGCCAGCCTGG 951
|||||
284 lyThrAspGlySerHisTyrTrpSerLysAsnTrpAlaLysAlaAla 300
|||||
952 GCACGTGATGCTCACAATCTGAGCAAGCACTGGGCCAAGGCTGCAGCC 1001
|||||
301 PheValThrSerProProValAsnProAspProThrThrAlaAspHisLe 317
|||||
1002 TTTGTCACTTCAACCCCTATCAACCCGCGCCCTACTACTGCAGACCACT 1051
|||||
317 uThrCysLeuLeuSerSerGlyAspLeuGlnValThrGlySerAlaHisC 334
|||||
1052 CACCTCTCTGCTGTCAGTGGGACCTCCAGGTGACAGGCAGTGCCTCACT 1101
|||||
334 yThrPheThrThrAlaGlnLysAlaValGlyLysAspAsnPhAlaLeu 350
|||||
1102 GCACCTTCACTACTGCCAAGGCTGTTGGCAAGACAACTTCACACTG 1151
|||||
351 IleProGluGlyThrAsnGlyIleGluGluArgMetSerMetValTrpGl 367
|||||
1152 ATCCCCGAGGTAGTCAACGGGTATAGAAGAGCGCATGTGTGTCTGGGA 1201
|||||
367 uLysCysValAlaSerGlyLysMetAspGluAsnGluPheValAlaValT 384
|||||
1202 GAATGTGTGCTTCAGGGAATGGAGAGATGAGTTCGTTGCCGGA 1251
|||||
384 hrSerThrAsnAlaAlaLysIlePheAsnPheTyrProArgLysGlyArg 400
|||||
1252 CCAGCAAAATGCTGCCAAAATCTCAATTTTACCCAGGAGGGCGGT 1301
|||||
401 ValAlaValGlySerAspAlaAspLeuValIleTrpAsnProLysAlaTh 417
|||||
1302 GTGGCCGTGGGCTGTGTGTGCTGAGCTGCTATCTGGAACCCAGGCGCAC 1351
|||||
417 rLysIleLeuSerAlaLysThrHisAsnLeuAsnValGluTyrAsnIleP 434
|||||
1352 GAATGTCATCTCTCCAGAGCCATAACCTGAATGTAGATACAACTCT 1401
|||||
434 heGluGlyValGluCysArgGlyAlaProAlaValIleSerGlnGly 450
|||||
1402 TTGAAGGAGTGGAGTGGCGAGGAGTGGCCACGGTGTATAGTCAGGGC 1451
|||||
451 ArgValAlaLeuGluAspGlyLysMetPheValThrProGlyAlaGlyAr 467
|||||
1452 AGAGTGTGTCTGGAGGAGGGAACCTGCTGTCTCATCCAGGGGTGGCGC 1501
|||||
467 gPheValProArgLysThrPheProAspPheValTyrLysArgIleLysA 484
|||||
1502 CTTTATTTCCCGGAGAGCTTCCCGGACTTGTCTATAGAGGATAAGG 1551
|||||
484 laArgAsnArgLeuAlaGluIleHisGlyValProArgGlyLeuTyrAsp 500
|||||
1552 CTCGCAACAGGCTAGCAGAGATCCACGGTGTGCTCGAGGCTGTACGAC 1601
|||||
501 GlyProValHisGluValMetValProAlaLysProGlySerGlyAlaPr 517
|||||
1602 GGGCTGTGCTGATGATGATGTTTACCTGCCAAGCAGGAACTGGCCACA 1651
|||||
517 oAlaArgAlaSerCysProGlyLysIleSerValProProValArgAsnL 534
|||||
1652 GGCCGTCATCTCTGTCAGCAAGATCTCAGTGCACCCCGTGGCCACACC 1701
|||||
534 euHisGlnSerGlyPheSerLeuSerGlySerGlnAlaAspAspHisIle 550
|||||

```

```

1702 TGCACAGTCGGGGTTCAGCCTATCTGCTCTCAGGCTGACGATCACATT 1751
|||||
551 AlaArgArgThrAlaGlnLysIleMetAlaProProGlyGlyArgSerAs 567
|||||
1752 GCCAGACGTACGGCTCAGAAGATCATGCACCCCCCGGAGGACGCTCCAA 1801
|||||
567 nileThrSerLeuSer 572
|||||
1802 CATCAGTCTCTTCC 1817
|||||
seq_name: gb_ro2:RNU52103
seq_documentation_block:
LOCUS RNU52103 2489 bp mRNA ROD 02-JUL-1996
DEFINITION Rattus norvegicus rCRMP-3 mRNA, partial cds.
ACCESSION U52103
VERSION U52103.1 GI:1399539
KEYWORDS
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 2489)
AUTHORS Wang, L. and Strittmatter, S.M.
TITLE A family of CRMP genes is differentially expressed in the nervous
system
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2489)
AUTHORS Strittmatter, S.M.
TITLE Direct Submission
JOURNAL Submitted (22-MAR-1996) Stephen M. Strittmatter, Yale
University School of Medicine, 333 Cedar Street, New Haven, CT
06520, USA
FEATURES
source
1..2489
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/tissue_type="brain"
/dev_stage="adult"
<1..1696
/note="similar to 62 kDa chick collapsin response mediator
protein"
/codon_start=2
/product="rCRMP-3"
/protein_id="AA03281.1"
/db_xref="GI:1399540"
/translation="IPRITSDRLIKGKIVNDQSFADLYVEDGLIKQIGENLIYP
GGIKTIDAHGLMVLPGGDVHTRLQMPVLGTPADDFCQGTAKALAGGTMILDHVP
DAGVSLIAAYEOWRERADSAACDYLHVDI PRMHTSKEELEALVRDKGNSFLVM
AYKDRCOCTDGOIYEIIFSLRDLGAVAQVHAENGDI VEEOKRLLEGQITGEGHLS
HPBEVEAEAVYRAVTAKQANCLYITKVMKGAADMAQAKRGGVVFGEPIITASLG
TDGSHYWSKNWAKAAAFVTPPINPDTTADHLTSLSSGDLQVTSAGHCTFTTAOKA
VGKDNFTLPIEGINGIEERMSVWKEKVCAGKMDENEFVAVTSTNAAKIFNFPKRG
LVVSGDADLVINPRATKVIYSAKSHNLNRYNIEFEGECRGLPTVVISQGRVLEDGN
LLVTPGAGREIPRTFPDFVYKRIKARNLAEIHGVPRLYDGPVHEVMLPAKPGSGT
QARASCPKISVPPVRNLHQSGLSGSQADDDHARRTAOKIMAPPGGRSNTLSL"
BASE COUNT 557 a 692 c 708 g 532 t
ORIGIN
alignment_scores:
Quality: 2789.00 Length: 564
Ratio: 5.025 Gaps: 0
Percent Similarity: 98.404 Percent Identity: 93.262
alignment_block:
US-09-367-496-8 x RNU52103
Align seg 1/1 to: RNU52103 from: 1 to: 2489
9 llaProArgIleThrSerAspArgLeuLeuIleArgGlyGlyArgIleVa 25
|||||

```

2 ATTCCCGGATACGAGTACCGCCCTTCTCATCAAGAGTGGGAAGATTCT 51
 25 lAsnAspGlnSerPheTyrAlaAspValHisValGluAspGlyLeuI 42
 52 GAATGATGACCACTCTTTCATCGCGATCTCTACGTGGAGGACGGTCTGA 101
 42 leIysGlnIleGlyGluAsnLeuIleValProGlyGlyIleHisThrIle 58
 102 TTAACAAATGGAGAAATCTTATCGTCCCTGGGGAATCAAAACCATC 151
 59 AspAlaHisGlyLeuMetValLeuProGlyGlyValAlaSpValHisThrAr 75
 152 GATGCTCATGGTCTGATGATGCTGCTGGGGAGTTGACGTCCACACCGG 201
 75 gLeuGlnMetProValLeuGlyMetThrProAlaAspPheCysGlnG 92
 202 GCTGCAGATGCTGCTGGCATGACCCACGCGATGATTCTGTCAGG 251
 92 lYThrLysAlaAlaLeuAlaGlyIleThrThrMetIleLeuAspHisVal 108
 252 GCACCAAGGCTCCCTAGCAGCGGGACCCACGATGATATTGACCATGTG 301
 109 PheProAspThrGlyValSerLeuLeuAlaAlaTyrGluGlnTrpArgGl 125
 302 TTTCTGATGCTGGCGTGAGCTGCTGGCGGCTATGACGATGGCGGA 351
 125 uArgAlaAspSerAlaAlaCysAspTyrSerLeuHisValAspIleT 142
 352 GCGAGCAGACAGTGCAGCGCTGCTGTGACTACTCTTACACGTGGACATTC 401
 142 hArgTrpHisGluSerIleLysGluGluLeuGluAlaLeuValLysGlu 158
 402 CCCGCTGGCATGAGACACCAAGAGGAGCTGGAGGCCCTAGTCAGGAC 451
 159 LysGlyValAsnSerPheLeuValPheMetAlaTyrLysAspArgCysGl 175
 452 AAAGTGTGAACCTCTCTGCTGCTTCATGGCTTACAAGCAGAGTGCCA 501
 175 nCysSerAspSerGlnMetTyrGluIlePheSerIleIleArgAspLeuG 192
 502 GTGTACTGATGGCCAGATATATGAATCTTCAGCCTCATCCGGGACCTGG 551
 192 lYAlaLeuAlaGlnValHisAlaGluAsnGlyAspIleValGluGlu 208
 552 GAGCTGTGGCCCGAGTGTGAGAAAACGGGGACATCGTGGAGAGGAA 601
 209 GlnLysArgLeuLeuGluLeuGlyIleThrGlyProGluGlyHisValIle 225
 602 CAGAAAGCCCTGCTGGACAGGCGATCACTGGTCTGAGGGCCATGTGCT 651
 225 uSerHisProGluGluValGluAlaGluAlaValTyrArgAlaValThrI 242
 652 CAGCACCCAGAAGAGGTAGAGGGCGGAAGCTGTGTACAGAGCTGTGACCA 701
 242 leAlaLysGlnAlaAsnCysProLeuTyrValThrLysValMetSerLys 258
 702 TTGCCAACAGGCCAACCTGCCCATCTGTATATACCAAGGTGATGAGCAAG 751
 259 GlyAlaAlaAspAlaIleAlaGlnAlaLysArgArgGlyValValValph 275
 752 GGTGCGGTGATATGTTGTTGCCAACGCCAAGCGCAGGGGGGTGGTCTCT 801
 275 eGlyGluProIleThrAlaSerLeuGlyThrAspGlySerHisTyrTrps 292
 802 TGGGGAACCTATCACTGCCACGCTGGGCATCTGATGGCTCACACTACTGGA 851
 292 erLysAsnTrpAlaLysAlaAlaPheValThrSerProProValAsn 308
 852 GCAGAACTGGGCCAAGGCTGCACCTTCGTCACATCACCCCTATCAAC 901
 309 ProAspProThrThrAlaAspHisLeuThrCysLeuLeuSerSerGlyAs 325
 902 CCGGACCCCTACTACTGACGACCACTACCTCTCTGCTGTCCAGTGGGA 951

325 pLeuGlnValThrGlySerAlaHisCysThrPheThrThrAlaGlnLysA 342
 952 CCTCCAGGTGCAGGCGAGTGCACACTGCACCTTCACTACTGCCCAAGG 1001
 342 laValGlyLysAspAsnPheAlaLeuIleProGluGlyThrAsnGlyIle 358
 1002 CTGTGGCAAGATAACTTCAAGCTGATCCCGAGGGCATCAACGGTATA 1051
 359 GluGluArgMetSerMetValTrpGluLysCysValAlaSerGlyLysMe 375
 1052 GAAGAACGCATGCTGTGTGCTGGGAAATGTGTGGCTTCAGGGAAT 1101
 375 tAspGluAsnGluPheValAlaValThrSerThrAsnAlaAlaLysIleP 392
 1102 GCACGAGATGAGTTCGTTCGCGTGCACGACCAAAATGCTGCCAAATCT 1151
 392 heAsnPheTyrProArgLysGlyArgValAlaValGlySerAspAlaAsp 408
 1152 TCAATTTCTACCCAGGAAGGGCGTGTGGCTGTGGCTCTGATGCTGAC 1201
 409 LeuValIleTrpAsnProLysAlaThrLysIleIleSerAlaLysThrHi 425
 1202 CTGGTCACTCTGAACCCCGAGGCCACAAAGTCACTCTGCCAAGAGCCA 1251
 425 sAsnLeuAsnValGluTyrAsnIlePheGluGlyValGluCysArgGlyA 442
 1252 TAACCTGAATGTAGAGTACAACATCTTCGAAGAGTGGAGTGCAGAGGCA 1301
 442 laProAlaValIleSerGlnGlyArgValAlaLeuGluAspGlyLys 458
 1302 TGCCACGCTGCTCAAGTACAGGCGAGAGTGTGTGGAGGACGGAAAC 1351
 459 MetPheValThrProGlyAlaGlyArgPheValProArgLysThrPhePr 475
 1352 CTGCTTGTCACTCCAGGGGCTGGCGCTTCATTCCCGGAAGACATTCCC 1401
 475 oAspPheValTyrLysArgIleLysAlaAlaArgAsnArgLeuAlaGluIleH 492
 1402 AGACTTGTCTATAGAGGATAAAGGCTCGTAACAGGCTAGCAGAGATCC 1451
 492 lSgYValProArgGlyLeuTyrAspGlyProValHisGluValMetVal 508
 1452 ACGGTGTGCCCGAGCGCTGTATGATGGCGCTGTCCAGAAAGTGATTA 1501
 509 ProAlaLysProGlySerGlyAlaProAlaArgAlaSerCysProGlyLy 525
 1502 CTGCCAAGCCAGGGAGCGGCACACAAAGCCCGTGTCTGTCCAGGCAA 1551
 525 sIleSerValProValArgAsnLeuHisGlnSerGlyPheSerLeus 542
 1552 GATCTCAGTGCACCCGTCGCAACCTGCACCAAGTCAGGGTTCAGCCTAT 1601
 542 erGlySerGlnAlaAspAspHisIleAlaArgArgThrAlaGlnLysIle 558
 1602 CTGGCTCTCAGGCTGACGATCATTGCGCAGACGACGCTCAGAAGATC 1651
 559 MetaLaProProGlyGlyArgSerAsnIleThrSerLeuSer 572
 1652 ATGGCACCCCGGAGGACGCTCCAAATCACTTCACTTCTTCT 1693

seq_name: gb_ov: AF249294

seq_documentation_block: 2470 bp mRNA VRT 06-DEC-2000

LOCUS AF249294 Gallus gallus collapsin response mediator protein 3 (CRMP3) mRNA,

DEFINITION complete cds.

ACCESSION AF249294

VERSION AF249294.1 GI:11559529

KEYWORDS chicken.

SOURCE Gallus gallus

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

REFERENCE
AUTHORS      Fukuda,M., Watakabe,I., Yuasa-Kawada,J., Kawachi,H., Kurolwa,A.,
              Matsuda,Y. and Noda,M.
TITLE        Molecular Characterization of CRMP5, a Novel Member of the
              Collapsin Response Mediator Protein Family
JOURNAL      J. Biol. Chem. 275 (48), 37957-37965 (2000)
PUBMED       10956643
REFERENCE     2 (bases 1 to 2470)
              Yuasa-Kawada,J., Watakabe,I. and Noda,M.
              Direct Submission
              Submitted (27-MAR-2000) Division of Molecular Neurobiology,
              National Institute for Basic Biology, Myodaiji-cho, Okazaki, Aichi
              444-8585, Japan
FEATURES     Location/Qualifiers
              1..2470
                /organism="Gallus gallus"
                /db_xref="taxon:9031"
                /tissue_type="retina"
                /dev_stage="embryonic day 8"
                /note="breed: White Leghorn"
              1..2470
                /gene="CRMP3"
                /db_xref="GeneID:14738"
                /gene="CRMP3"
                /codon_start=1
                /product="collapsin response mediator protein 3"
                /protein_id="AA037997.1"
                /db_xref="GI:11599530"
                /translation="MSPNPGSGRGTSEQLIKGKIIVNDQSFYADIYEDGLIKQ
              IGENLVPSGVRTVDAYGOLVPGGIDVTRQLQVAVMGMSADDFYOGTRALAGGT
              MIMDLVPEAGTSLIAAYEQWRQADSRACCDYALHIDIPRHESLELEALVKDKG
              VNSFLVWAYKDLQCTDAOMYEIFCIIRDLGALAQVHAENGDIIEEOKRLDIDGIT
              GPEGVLSRPEVEAEAVYRAITIAKQANCPLYVTKIMSKSAADVAAQAKRGTVVYG
              EPTASLGDSHVSWSKWKAKAAFTVSPPLSPDPTTPERLSLLSCDLQVAGSAHC
              TPTAAKAGKNDTLLIPEGTNGIEERMAIIWEKCVVSGKMDNDFVAVTSTNAKIF
              NCYPRKGRVIAVSDADLVLMNPRAIKVISAKTHNLNVEYNIFEGTECHGAPAVVISOG
              KVLVEDGLNLFYTESGSEVPRKTPDFYVKKIKARNLEAEVHGVPRLGIDGPFVHDVIL
              STKAVPTTIAASRIACAGKVPVPRNLHQSGFSLSGSQADHDHVARRTAQKIMAPPGG
              RSNITSLs"
BASE COUNT   533 a 685 c 757 g 495 t
ORIGIN
alignment_scores:
  Quality: 2517.00      Length: 574
  Ratio: 4.618          Gaps: 1
  Percent Similarity: 94.948      Percent Identity: 81.359
alignment_block:
  US-09-367-496-8 x AF249294 ..
Align seg 1/1 to: AF249294 from: 1 to: 2470
1 MetSerPheGlnGlyLysLysSerIleProArgIleThrSerAspArgLe 17
|||||
14 ATGTGCAACCGGCGAGAGAGCGGCCCGCGGACGAGCGAGCGAGCT 63
|||||
17 uleuIleArgGlyGlyArgIleValAsnAspGlnSerPheTyrAlaA 34
|||||
64 GCTGATCAAGGGGGGAAATCGTCACGACGACGACGACGATGCGT 113
|||||
34 spValHisValGluAspGlyLeuIleLysGlnIleGlyGluAsnLeu 50
|||||
114 ACATTTATGTGGAAGATGGTGTGATAAACACAGATTGGAGAACCT 163
|||||
51 ValProGlyGlyIleHisThrIleAspAlaHisGlyLeuMetValLeu 67
|||||
164 GTTCCACGGGGGTGAGGAGCGGTGATGCCACGGGACGCTGGTGTC 213
|||||
67 oGlyGlyValAspValHisThrArgLeuGlnMetProValLeuGlyMet 84
|||||

```

```

214 GGGGGGCGATCGATGTCACACGGCGCTGCAGATGCGCTGTGTGATGGGATGG 263
84 hrProAlaAspAspPheCysGlnGlyThrLysAlaAlaLeuAlaGlyGly 100
:: |||||
264 CCTCTGCTGATGACTTCTACCAAGGCACGAGGCGCCCTGGCGGGAGGA 313
|||||
101 ThrThrMetIleLeuAspHisValPheProAspThrGlyValSerLeuLe 117
|||||
314 ACCACGATGATAATAGGACCATGCTGCTGCCGAGGCTGGGAGGACCTGCT 363
|||||
117 uAlaAlaTyrGluGlnTrpArgGluArgAlaAspSerAlaAlaCysCysA 134
|||||
364 GGCCTGCTACGAGCAGTGGCGCAGCGTGTGACAGCAGGCGCTGCTGCG 413
|||||
134 spTyrSerLeuHisValAspIleThrArgTrpHisGluSerIleLysGlu 150
|||||
414 ACTAGCGCTGCACATGACATCCCGCTGGCATGAGAGCTGCGGGAG 463
|||||
151 GluLeuGluAlaLeuValLysGluLysGlyValAsnSerPheLeuValPh 167
|||||
464 GAGCTGGAGGCGCTGGTCAAGACACAGGCTGTGAACCTCTCTCTGGTTT 513
|||||
167 eMetAlaTyrLysAspArgCysGlnCysSerAspSerGlnMetTyrGluI 184
|||||
514 TATGCGCTACAAAGGACAGCTGACGTGACCGACCGCCAGATGTATGAA 563
|||||
184 IePheSerIleIleArgAspLeuGlyAlaLeuAlaGlnValHisAlaGlu 200
|||||
564 TATTCTGCTATTATCGACCTTGGGGCCATAGCCCAAGTGCACCGCTGAA 613
|||||
201 AsnGlyAspIleValIleGluGlnLysArgLeuLeuGluLeuGlyIyl 217
|||||
614 AACGGGACATCATCGAGGAGGAGCAGAGAGGCTGCTGGACATTTGGAT 663
|||||
217 eThrGlyProGluGlyHisValLeuSerHisProGluGluValGluAlaG 234
|||||
664 CACTGGGCGAGAGGCGGACGTCTGAGCGGCTGAGGAGGTTGAAGCAG 713
|||||
234 luAlaValTyrArgAlaValThrIleAlaLysGlnAlaAsnCysProLeu 250
|||||
714 AGCGGTGTACCTGCAATACATAGCAAAACAAGCCAACTGCCCTTG 763
|||||
251 TyrValThrLysValMetSerLysGlyAlaAlaAspAlaIleAlaGlnAl 267
|||||
764 TAGCTCACCACAAATAATGACAAAAGTGCAGCAGATGTGGTAGCTCAAGC 813
|||||
267 aLysArgArgGlyValValValPheGlyGluProIleThrAlaSerLeuG 284
|||||
814 AAAGAGAAAGGACCTGTGTGTGTCGGGAGGCCCATCATCTGCCAGCCTG 863
|||||
284 lYThrAspGlySerHisTyrTrpSerLysAsnTrpAlaLysAlaAlaA 300
|||||
864 GTGCTGACGCGCTCGCACTACTGAGCAAGAACTGGGCAAGGCTGCAGCC 913
|||||
301 PheValThrSerProProValAsnProAspProThrThrAlaAspHisLe 317
|||||
914 TTGCTCACCTCACCGCGCTCAGCCCGGACCCACCCAGCCAGAGCGCCT 963
|||||
317 uThrCysLeuLeuSerSerGlyAspLeuGlnValThrGlySerAlaHisC 334
|||||
964 CTGCTCCCTGCTGCTGCTGGGACCTGTCAGGTGGCAGGAGTGTCTACT 1013
|||||
334 ysThrPheThrThrAlaGlnLysAlaValGlyLysAspAsnPheAlaLeu 350
|||||
1014 GCAGTTTCACTACCGCCCGCAGAGGCGCTGGGAAGGACAACTTCACCCTC 1063
|||||
351 IleProGluGlyThrAsnGlyIleGluGluArgMetSerMetValTrpGl 367
|||||
1064 ATCCCCGAGGGGAGAAATGTCATCGAGGCGGATGGCCATAATCTGGGA 1113
|||||
367 uLysCysValAlaSerGlyLysMetAspGluAsnGluPheValAlaValT 384
|||||
1114 GAAGTGGCTGCTCTCAGGGAAGATGGATGAGACGATTCGTGGCGGTGA 1163

```

```

384 hrSerThrAsnAlaAlaLysIlePheAsnPheThrProArgLysGlyArg 400
|||||
1164 CCAGCACTAATGCTGCCAAATCTTTAACTGCTACCCAGGAGGGCGA 1213
|||||
401 ValAlaValGlySerAspAlaAspLeuValIleThrPAsnProLysAla 417
|||||
1214 ATCGCGGGTGGCTGTGATGCGAGCTGTGCTGTGTGGAACCCAGGCTAC 1263
|||||
417 rLysIleIleSerAlaLysThrHisAsnLeuAsnValGluThrAsnIleP 434
|||||
1264 AAAAGTCATCTCGCAAAACCCACCAATTTGAATGTGGAGTACACATAT 1313
|||||
434 heGluGlyValGluCysArgGlyAlaProAlaValValIleSerGlnGly 450
|||||
1314 TTGAAGGACGAGGTGTACGCGGCTCTGCTGTGTGTCATCAGTCAGGCG 1363
|||||
451 ArgValAlaLeuGluAspGlyLysMetPheValThrProGlyAlaGlyAr 467
|||||
1364 AAAGTCGTCTTGAGGACGGGAACCTGTTGTCCACCGAGGCTCGGGCG 1413
|||||
467 gPheValProArgLysThrPheProAspPheValTyrLysArgIleLysA 484
|||||
1414 CTTCGTTCCCAAGAAAGAGCTTCCCGGATTTGTTTATAAAGAATAAAGG 1463
|||||
484 laArgAsnArgLeuAlaGluIleHisGlyValProArgGlyLeuTyrAsp 500
|||||
1464 CGAGGAACAGCTGGCGGAGGTGCACGGTGTCCACCGGGGCTGTATGAC 1513
|||||
501 GlyProValHisGluValMetValProAlaLys.....ProGlySerGI 515
|||||
1514 GGCCGCTCCACGATGTCGTCCTCAGCACCAAGCTGTGCCACCATCGC 1563
|||||
515 yAlaProAlaArgAlaSerCysProGlyLysIleSerValProValA 532
|||||
1564 GGCTCCAGGATCGCGCTGCGCGGCAAGTCCCGCGTCCCGTGC 1613
|||||
532 rGAsnLeuHisGlnSerGlyPheSerLeuSerGlySerGlnAlaAspAsp 548
|||||
1614 GCAACCTGCACAGTGGGGTTCAGCTGTCCAGGCTGTACAGGTGACGAC 1663
|||||
549 HisIleAlaArgArgThrAlaGlnLysIleMetAlaProProGlyLys 565
|||||
1664 CAGGTTGCCCGGCGCACAGCCAGAGATCATGGCCCCCGCGCGCG 1713
|||||
565 gSerAsnIleThrSerLeuSer 572
|||||
1714 CTCACACATCACTCCCTGTCC 1735
|||||

```

seq_name: gb_ov:GGU17277

```

seq_documentation_block:
LOCUS GGU17277 2423 bp mRNA VRT 02-APR-1996
DEFINITION Gallus gallus collapsin response mediator protein CRMP-62 mRNA,
complete cds.
ACCESSION U17277
VERSION U17277.1 GI:882146
KEYWORDS chicken.
SOURCE Gallus gallus
ORGANISM Gallus gallus
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
1 (bases 1 to 2423)
Goshima,Y., Nakamura,F., Strittmatter,P. and Strittmatter,S.M.
Collapsin-induced growth cone collapse mediated by an intracellular
protein related to UNC-33
Nature 376 (6540), 509-514 (1995)
95364923
2 (bases 1 to 2423)
Strittmatter,S.M.
AUTHORS Strittmatter,S.M.
TITLE Direct Submission
JOURNAL Submitted (17-NOV-1994) Stephen M. Strittmatter, Neurology, Yale

```

```

University School of Medicine, 333 Cedar Street, New Haven, CT
06520, USA
FEATURES
source
1. .2423
Location/Qualifiers
/organism="Gallus gallus"
/db_xref="taxon:9031"
/tissue_type="dorsal root ganglion"
/dev_stage="embryonic day 7"
302. .2020
/note="62 kDa collapsin response mediator protein"
/codon_start=1
/product="CRMP-62"
/protein_id="AAA93200.1"
/db_xref="GI:882147"
/translation="MSYQKGMVIFGGIDVHTRFOMPEQMTSADDFQGTAAALAGGTT
IGENLIVPGVKTI EAHRMVI PGGIDVHTRFOMPEQMTSADDFQGTAAALAGGTT
MIIDHVVPFGVTSLLTAFQDREWADSKCCDYSLAYDITEHKGQVEAEALKDHG
VNSFLVYMAFKDFEQLSDSQIYEVLSVIRDIGATAQVHAENGDI IAEEOQRILELGIT
GPEGHVLSRPEVEAEAVNAITIAQTNCPLYITKVMKSAAEVIQAQRKGTVVYG
EPITASLDGDSHYNSKNWAKAAAFVTSPLSPDPTPDFLNSLLSCGDLQVTSAGHC
TENTAKAVGKDNFTLIPETNGTEERMSI IWDKAVTGMKDNQFVAVTSTNAATF
NLYPRKGRIAGVSDADLVIPDPSVKTISAKTHNISLEYNIFEGMECRGSLVVISQG
KIVLEDGNLHVTEGSGRYIPRKPFDPFVVKRIKARSLEALRGVPRGLYDGPVCEVSV
TPKTVTTPASSAKTSPAKQAPPVPRNLHQSGFSLSGAQIDIDNIPRRTTQRIVAPPGGRA
NITSLG"
BASE COUNT 593 a 613 c 654 g 563 t
ORIGIN

```

alignment_scores:

Quality: 2362.00 Length: 571
Ratio: 4.457 Gaps: 0
Percent Similarity: 92.820 Percent Identity: 76.357

alignment_block:

US-09-367-496-8 x GGU17277

Align seg 1/1 to: GGU17277 from: 1 to: 2423

```

1 MetSerPheGlnGlyLysSerIleProArgIleThrSerAspArgLe 17
|||||
302 ATGCTTTACCGGGGAGAAAGAACATCCCGCGGATCACGACGATCGTCT 351
|||||
17 uLeuIleArgGlyGlyArgIleValAsnAspAspGlnSerPheTyrAla 34
|||||
352 TCTGATCAAGGTGGGAGATCGTCAATGACGACCATCTTCTTATGCAG 401
|||||
34 spValHisValGluAspGlyLeuIleLysGlnIleGlyGluAsnLeuIle 50
|||||
402 ACATCTACATGGAAGATGGGCTCATCAAGCAATAGGGAGATCTCATC 451
|||||
51 ValProGlyGlyIleHisThrIleAspAlaHisGlyLeuMetValLeuPr 67
|||||
452 GTGCCCGGAGGAGTGAACACCATCGAAGCCACGCGCAGGATGGTGATCC 501
|||||
67 OGlyGlyValAspValHisThrArgLeuGlnMetProValLeuGlyMet 84.
|||||
502 TGGAGGGATCGATGTTCAACCGCTTCCAGATGCCAGACGAGGGATGA 551
|||||
84 hrProAlaAspAspPheCysGlnGlyThrLysAlaAlaLeuAlaGlyGly 100
|||||
552 CTCTCGCTGATCACTTCTTCCAAAGGAGACCAAGGCTCGCTGGTGGGC 601
|||||
101 ThrThrMetIleLeuAspHisValPheProAspThrGlyValSerLeuLe 117
|||||
602 ACCACCATGATCATGATCATGTGGTCTCTGAGCCCGGACGAGTTTGT 651
|||||
117 uAlaAlaTyrGluGlnTrpArgGluAlaAspSerAlaAlaCysCysA 134
|||||
652 GACTGCATTTGACCACTGAGAGATGGCAGACACAAATCGTCTGTG 701
|||||
134 spTyrSerLeuHisValAspIleThrArgTrpHisGluSerIleLysGlu 150
|||||

```


Tue Jul 31 13:08:18 2001

702 ATTACTCCTCGCAGCTGGACATCCAGGAGTGCGACAAAGGGCTCCAGGAG 751
151 GluLeuGluAlaLeuValValGlyGluValAlaAsnSerPheLeuValph 167
752 GAGATGGAAGCTCTGGTTAAAGATCAGCGGTGGAATTCTATCTCTGGTTA 801
167 eMetaLalValLysAspArgCysGlnCysSerAspSerGlnMetTyrGluI 184
802 TATGGCTTCAAGACCGCTTCCAGCTGCTGACGTCGAGATCTATGAGG 851
184 lePheSerIleleArgAspLeuGlyAlaLeuAlaGlnValHisAlaGlu 200
852 TCCTGAGTGTATCCGGGATATGTTGGCACAGCTCAAGTGCATGCTGAG 901
201 AsnGlyAspIleValGluGluGlnLysArgLeuLeuGluLeuGlyI 217
902 ATGTGTGACATCATTCCTCCGAGGAGCAGCAAGATCTTGGAGCTGGAAT 951
217 eThrGlyProGluGlyHisValLeuSerHisProGluGluValGluAlaG 234
952 CACAGCCCTGAAGGCGACGTGTTGAGCAGACCTGAAGAGGTGGAGGCG 1001
234 luAlaValTyrArgAlaValThrIleAlaLysGlnAlaAsnCysProLeu 250
1002 AAGCTGTGAACCGGGCAATCACCATCGCCACCAACCAACTGCCCTCTC 1051
251 TyrValThrLysValMetSerLysGlyAlaAlaAspAlaIleAlaGln 267
1052 TATATCACCAAGGTGATGAGCAAAAGCGTGTGAAGTCATTTGCTCAAGC 1101
267 aLysArgGlyValValValPheGlyGluProIleThrAlaSerLeuG 284
1102 CAGGAAAGGCGACGTGTGTGTGGGAGGCCCATCACAGCAGCTTGG 1151
284 lyThrAspGlySerHisTyrTrpSerLysAsnTrpAlaLysAlaAla 300
1152 GTACTGATGATCTCATCTGAGCAGCAAGATTTGGGGAAGGCGCGCT 1201
301 PheValThrSerProProValAsnProAspProThrThrAlaAspHisLe 317
1202 TTTGTCACTTCCCAACCTGAGTGTGCTGACCAACCACTCTGATTTTCT 1251
317 uThrCysLeuLeuSerSerGlyAspLeuGlnValThrGlySerAlaHisC 334
1252 CAACCTCACTGCTGCTGCTGCGGGACCTCAAGTTACTTGGCAGTCCCACT 1301
334 ysThrPheThrAlaGlnLysAlaValGlyLysAspAsnPheAlaLeu 350
1302 GCACCTTCAACACTGCTCAGAAAGCTGTGGGAAGGACAACTTCAACCTA 1351
351 IleProGluGlyThrAsnGlyIleGluGluArgMetSerMetValTyrG 367
1352 ATCCCGAGGGGACCAAGCGGACTGAGGAGAGGATGCTCATCATCTGGGA 1401
367 uLysCysValAlaSerGlyLysMetAspGluAsnGluPheValAlaVal 384
1402 CAGCTGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1451
384 hrSerThrAsnAlaLysIlePheAsnPheTyrProArgGlyGlyArg 400
1452 CCAGCACAATGTCAGCTAAATCTTTAACTTGTATCCACGAGGAGGCGGT 1501
401 ValAlaValGlySerAspAlaAspLeuValIleTrpAsnProLysAlaTh 417
1502 ATTCAGTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1551
417 lLysIleIleSerAlaLysThrHisAsnLeuAsnValGluTyrAsnIleP 434
1552 CAAGACAATCTCTGCAAGACCCCAACATATCTCTGGAGTACAACATCT 1601
434 heGluGlyValGluCysArgGlyAlaProAlaValIleSerGlnGly 450
1602 TTGAAGGCAATGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1651

451 ArgValAlaLeuGluAspGlyLysMetPheValThrProGlyAlaGlyAr 467
1652 AAGATCGTGTGGAGGATGGAAATCTCCAGTAACGGAGGGGTCTGGAG 1701
467 gPheValProArgLysThrPheProAspPheValTyrLysArgIleLysA 484
1702 GTAATATCCCAAGAAACATTTCCCTGACTTCTGTTTACAAAGGCATCAAG 1751
484 laaAgAsnArgLeuAlaGluIleHisGlyValProArgGlyLeuTyrAsp 500
1752 CAAGGAGCAGGCTGCGGAGCTGCGGAGTGTGCTCGAGGTCTCTATGAT 1801
501 GlyProValHisGluValMetValProAlaLysProGlySerGlyAlaPr 517
1802 GGGCGGCTCTCGGAAGTCTCGGTGACGCCGAGAGCGGTCACTCTCGCATC 1851
517 oAlaArgAlaSerCysProGlyLysIleSerValProProValArgAsnL 534
1852 TTCAGCTAAACATCTCTGCCAAACAGCAAGCCCACTGTGAGGAACC 1901
534 euHisGlnSerGlyPheSerLeuSerGlySerGlnAlaAspHisIle 550
1902 TGCACCACTCTGGCTTACCTGTCTGGCGCACAGATCGACGACCAACATC 1951
551 AlaArgAlaThrAlaGlnLysIleMetAlaProProGlyGlyArgSerAs 567
1952 CCACGCCGACCCAGCGCATCTGCGCACCGCGCGGCGGACGTGCCAA 2001
567 nIleThrSerLeu 571
2002 CATCACCAGCTTG 2014
seq_name: gt_om: BTU83278
seq_documentation_block: 3884 bp DNA MAM 08-OCT-1997
LOCUS BTU83278 Bos taurus neural specific protein CRMP-2 gene, complete cds.
DEFINITION U83278
ACCESSION U83278
VERSION U83278.1 GI:1916226
KEYWORDS
SOURCE
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
REFERENCE
AUTHORS 1 (bases 1 to 3884)
TITLE Kamata, T.K.
Isolation and characterization of a neural specific protein (Nsp60)
gene homologous to C. elegans unc-33 controlling axonal outgrowth
and guidance
JOURNAL Brain Res. Mol. Brain Res. (1997) In press
REFERENCE 2 (bases 1 to 3884)
AUTHORS Kamata, T.K.
TITLE Direct Submission
JOURNAL Submitted (27-DEC-1996) Lab. of Biochemical Physiology, National
Cancer Inst. FCRDC, P.O. Box B, Bldg 567, Frederick, MD 21703, USA
FEATURES
SOURCE
1..3884
/organism="Bos taurus"
/db_xref="taxon:9913"
/note="map location 8p21 in human; mouse 14"
177..1895
/note="unc-33 homolog"
/codon_start=1
/product="neural specific protein CRMP-2"
/protein_id="AAB80618.1"
/db_xref="GI:1916227"
/translation="MSYQKKNIPRTSDRLIKGKIWNDDQSFVADYIMEDGLIKQ
IGENLIVPGGVKTIHAHSRWIPGGIDVHTRFQMPDQGTADDFFQCTKAALAGTT
MIDHVVPEPTSLAAFDQWREWDKSCDDYSLRVDITEWHKGVQBEKEALVYDHG
VNSILVYMAFKDRFQDTQSIYEVLSVIRDIGAIVHAENGDIIEQQRILDIGIT
GPEGHVLSRPEVEAEVAVNRKITIANQTNCLPIYTKVMSAAAEVIAQARKKGTYYVG

[illegible]

BASE COUNT	NITS LG"		BASE COUNT	others
ORIGIN	911 a	1031 c	1029 g	912 t

alignment_scores:

Quality:	2347.00	Length:	571
Ratio:	4.437	Gaps:	0
Percent Similarity:	92.644	Percent Identity:	75.657

alignment_block:

US-09-367-496-8 x BTU83278

Align seg 1/1 to: BTU83278 from: 1 to: 3884

1 MetSerPheGlnGlyLysLysSerIleProArgIleThrSerAspArgLe 17
177 ATGTCTTATCAGGGGAAGAAAATAATTCACGCATCACGAGTGTCT 226
17 uEuIleAtrgGlyGlyArgIleValAsnAspAspGlnSerPheTyrAlaA 34
227 TCTGATCAAGGAGGTAAATGTTTAAATGATGACCAAGTCTTCTTATGCAG 276
34 spValHisValGluAspGlyLeuIleLysGlnIleGlyGluAsnLeuIle 50
277 ACATATACATGGAAGATGGTTCATCAACCAATAGGAGAAAACCTGATT 326
51 ValProGlyGlyIleHisThrIleAspAlaHisGlyLeuMetValLeuPr 67
327 GTCCCGGGAGGGGTGAAGACCAATTGAAGGCCAATTCAGAAATGGTGATCCC 376
67 ogGlyGlyValAspValHisThrArgLeuGlnMetProValLeuGlyMetT 84
377 TGGAGGATTGACGTCCATTACCCGTTTCAGATGCTGATCAGGGGATGA 426
84 hrProAlaAspPheCysGlnGlyThrLysAlaAlaLeuAlaGlyGly 100
427 CCTCTGCTGATCACTTCTTCAAGGCACCAAGGCAGCCCTGGCTGGAGGA 476
101 ThrThrMetIleLeuAspHisValPheProAspThrGlyValSerLeuLe 117
477 ACCACATGATCATTTGACACAGTGTTCTTGAGCCTGGGACAAAGCTGCT 526
117 uAlaAlaTyrGluGlnTrrArgGluArgAlaAspSerAlaAlaCysCysA 134
527 GSCCGCCTTTGACCAGTGGCGGAATGGCTGACAGCAAGTCCTGCTGTG 576
134 spTyrSerLeuHisValAspIleThrArgTrpHisGluSerIleLysGlu 150
577 ACTACTCTCTGCAGTGTGACATCACTGAATGGCACAAAGGGGTCACGAA 626
151 GluLeuGluAlaLeuValLysGluLysGlyValAsnSerPheLeuValph 167
627 GAGATGGAGGCCCTGGTGAAGAGCACATGGGTGAATTCCTTCCTTTGTGA 676
167 eMetAlaTyrLysAspArgCysGlnCysSerAspSerGlnMetTyrGluI 184
677 CATGGCTTCAAAAGATCGTTCCAGCTCCAGGATTCACAGATCTATGCAAG 726
184 lePheSerIleAlaArgAspLeuGlyAlaLeuGlnValHisAlaGlu 200
727 TCTTGAGCGGTATCCGGATATTGGACGGATTCGCCAGGTCCACGCGGAA 776
201 AsnGlyAspIleValGluGluGlnLysArgLeuLeuGluLeuGlyI 217
777 AACGGCGACATCATCGCAGAGGACACACAGAGATCTCTGGATCTGGGCAT 826
217 eThrGlyProGluGlyHisValLeuSerHisProGluGluValGluAlaG 234

1727 CTGGCTAAGACGCTCTCCGCCAAGCAGCAGGCCACCTGTCGGGACCC 1776
 534 euHisGlnSerGlyPheSerLeuSerGlySerGlnAlaAspHisIle 550
 1777 TGCACCAAGTCTGGATTTCAGCTTCTCCGGTCCCGAGATTGATGACAATC 1826
 551 AlaArgArgThrAlaGlnLysIleMetAlaProGlyGlyArgSerAs 567
 1827 CCTCGCCGACACCGACGATCGTGGCGCCCTTGGTGGCGGTGCTAA 1876

567 nileThrSerLeu 571
 1877 CATCACCAAGCTG 1889

seq_name: gb_ro2:RRTOAD64

seq_documentation_block: 2947 bp mRNA ROD 23-JAN-1996

LOCUS RRTOAD64

DEFINITION R.rattus mRNA for T0AD-64.

ACCESSION Z46882

VERSION 246882.1 GI:599965

KEYWORDS TOAD-64.

SOURCE black rat.

ORGANISM Rattus rattus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

REFERENCE 1 (bases 1 to 2947)

AUTHORS Minturn,J.E., Fryer,H.J., Geschwind,D.H. and Hockfield,S.

TITLE TOAD-64, a gene expressed early in neuronal differentiation in the

rat. is related to unc-33, a C. elegans gene involved in axon

outgrowth

J. Neurosci. 15 (10), 6757-6766 (1995)

MEDLINE 96033765

REFERENCE 2 (bases 1 to 2947)

AUTHORS Hockfield,S.

TITLE Direct Submission

Submitted (14-NOV-1994) Susan Hockfield, Neurobiology, Yale

University School of Medicine, 333 Cedar Street, New Haven, CT,

06510, USA

FEATURES

source

1..2947

/organism="Rattus rattus"

/strain="Sprague-Dawley"

/db_xref="taxon:10117"

/clone="TOAD64 (turned on after 'division, 64kd)"

/tissue_type="brain"

/clone_lib="P5 rat pBluescript brain library"

178..1896

/codon_start=1

/product="TOAD-64"

/protein_id="CAA86981.1"

/db_xref="GI:599966"

/db_xref="SWISS-PROT:P47942"

/translation="MSYQKKNIPRITSDRLIKGGKIVNDQSFVADITYMEDGLIKQ

IGNLIVPGGVKTI EAHSRMVTPGGIDVTRFQMPDQMTSADDFQGTKAALAGGTT

MTDHWVPGGVKTI EAHSRMVTPGGIDVTRFQMPDQMTSADDFQGTKAALAGGTT

VNSFLVMAFKRQITDSQIYEVLSVIRDIGALQVHAENGDI AEEOQRIIDLGIT

GPBGHVLSPREEAEAVANRSITIANQNCPLFTVTKVMSKSAEVAQAQKKGTVYVG

EPITASLIGTDSGHYWSKNWAKAAAFVTSPLSPDPTTDFLNSLSLSCGLDQVTGSAHC

TFNTAQKAVGKDNFTLIPETNGTEERMSVINDKAVVTKMDENQFVAVTSTNAKVF

NLYPRKGRISVSDADLVITWDPDSVKTISAKTHNSALEYNIPEGMECRGSPLVVISQ

KIVLEDGTLHVTGSGRYIPRKPFDPFVKIKRSLRSLAEKRGVPRGLYDGPVCEYSV

TPKTVTPASSAKTSPAKQAPPVRLNHSQSFSLSGAIDIDNIPRRRTQRIVAPPGGRA

NITSLSG"

BASE COUNT 666 a 765 c 769 g 747 t

ORIGIN

alignment_scores:

Quality: 2343.00 Length: 571

Ratio: 4.438 Gaps: 0

Percent Similarity: 92.469 Percent Identity: 75.832

alignment_block:

US-09-367-496-8 x RRTOAD64

Align seg 1/1 to: RRTOAD64 from: 1 to: 2947

1 MetSerPheGlnGlyLysLysSerIleProArgIleThrSerAspArgLe 17
 178 ATCTCTTATCAGGGAAGAAATATTCACGCATCAGCAGCGATCGTCT 227
 17 uLeuIleArgGlyGlyArgIleValAsnAspAspGlnSerPheTyrAlaA 34
 228 TCTGATCAAAAGGTGGCAAGATTGTGAATGATGACCACTCTTCTATGCAG 277
 34 spValHisValGluAspGlyLeuIleLysGlnIleGlyGluAsnLeuIle 50
 278 ACATATACATGGAAGATGGTGTGATCAAGCAATAGGAGAAAACCTGATT 327
 51 ValProGlyGlyIleHisThrIleAspAlaHisGlyLeuMetValLeuPr 67
 328 GTGCCAGGAGGGGTGAAGACCATCGAAGCCCACTCCAGAAATGGTGATCCC 377
 67 oGlyGlyValAspValHisThrArgLeuGlnMetProValLeuGlyMetT 84
 378 TGAAGGAATTGAGTGCACACTCGCTTCCAGATGCCAGACAGGGAATGA 427
 84 hrProAlaAspPheCysGlnGlyThrLysAlaAlaLeuAlaGlyGly 100
 428 CATCAGCTGATGACTTCTTCCAGGGAACCAAGCAGCCCTTGGCCGAGGA 477
 101 ThrThrMetIleLeuAspHisValPheProAspThrGlyValSerLeuLe 117
 478 ACCACCATGATCATCGACCATGTTGTTCTGAGCCCGGACCAAGCCTATT 527
 117 uAaAlaTyrGluGlnTrpArgGluArgAlaAspSerAlaAlaCysCysA 134
 528 GGCAGAGCTTTGATCAGTGGAGGAGTGGCGGACAGCAAGTCTCTGTGTG 577
 134 spTyrSerLeuHisValAspIleThrArgTrpHisGluSerIleLysGlu 150
 578 ACATATCGCTGACGTGGACATCAGGAGTGGCACAAGGCGCATCCAGGAG 627
 151 GluLeuGluAlaLeuValLysGlyLysGlyValAsnSerPheLeuValPh 167
 628 GAGATGGAAGCTCTGTTGAAGGACCAAGCGGGTAAACTCTCTCTCGTGA 677
 167 eMetAlaTyrLysAspArgCysGlnCysSerAspSerGlnMetTyrGluI 184
 678 CATGGCTTTCAAAGATCGTTCCAGCTGCGGATTTCCAGATCTATGAAG 727
 184 lePheSerIleIleArgAspLeuGlyAlaLeuAlaGlnValHisAlaGlu 200
 728 TACTGACGCTGATCCGGGATATTTGTCCTAGTCAAGTCCATGCAGAG 777
 201 AsnGlyAspIleValGluGluGluGlnLysArgLeuLeuLeuGlyIle 217
 778 AATGCTGACATCATTTGACGAGGAACACAGCAGAGGATCTGATCTGGCAT 827
 217 eThrGlyProGluGlyHisValLeuSerHisProGluGluValGluAlaG 234
 828 CACAGGCCCCGAGGACAGCTGCTGACCGCCGACAGAGGATCGAGGCTG 877
 234 luAlaValTyrArgAlaValThrIleAlaLysGlnAlaAsnCysProLeu 250
 878 AA3CTGTGAACCGGTCTCATCCATTTGCCAATCAGACCAACTGCCCGCTG 927
 251 TyrValThrLysValMetSerLysGlyAlaAlaAspAlaIleAlaGlnAl 267
 928 TATGTCACCAAGGTGATGACCAAGAGTGTGCTGAAGTCATCGCCAGGC 977
 267 aLysArgArgGlyValValValPheGlyGluProIleThrAlaSerLeuLeu 284
 978 ACGGAAGAGGAACTGTGGTGTATGTGTGAGCCCATCACTCCAGCCTGG 1027

```

284 lyThrAspGlySerHisTyrTrpSerLysAsnTrpAlaLysAlaAla 300
|||||
1028 GGACTGATGGCTCTCATATTGGAGCAAGAACTGGCCAGGCGCTGCC 1077
301 PheValThrSerProProValAsnProAspProThrThrAlaAspHisLe 317
|||||
1078 TTTGTACCTCTCCACCTTTCAGCCCTGAGCCGCGCCACCACTCCAGACTTCT 1127
317 uThrCysLeuLeuSerSerGlyAspLeuGlnValThrGlySerAlaHisC 334
|||||
1128 CAACTCGTTCCTGCTGCTGGAGACCTCCAGGTCAGTCCAGTGGCCACT 1177
334 yThrPheThrThrAlaGlnLysAlaValGlyLysAspAsnPheAlaLeu 350
|||||
1178 GTACCTTCAACACTGCCAGAGGCTGTGGGAAGGATAACTTCACCTG 1227
351 IleProGluGlyThrAsnGlyIleGluGluArgMetSerMetValTrpGI 367
|||||
1228 ATCCAGAGGCGACCAATGACCTGAGGAGCGGATGCTGTCTATTGGGA 1277
367 uLysCysValAlaSerGlyLysMetAspGluAsnGluPheValAlaValT 384
|||||
1278 TAAAGCTGTGCTCACTGGGAAGATGGACGAGAACCACTTTGTGCTGA 1327
384 hrSerThrAsnAlaAlaLysLysIlePheAsnPheTyrProArgLysGlyArg 400
|||||
1328 CTAGCACCAACGCGAGCAAGTCTCAATCTTTACCCACGGAAGGTCGT 1377
401 ValAlaValGlySerAspAlaAspLeuValIleTrpAsnProLysAlaTh 417
|||||
1378 ATCTCGTGGATCTGACGACGACCTGGTGTGATCTGGGACCTGACAGTGT 1427
417 rLysIleLeuSerAlaLysThrHisAsnLeuAsnValGluTyrAsnIleP 434
|||||
1428 GAAGACCATCTGTGCCAAGACGACCAACAGTCTCTTGTAGTACACATCT 1477
434 heGluGlyValGluCysArgGlyAlaProAlaValValIleSerGlnGly 450
|||||
1478 TTGAAGCATGAGTGTGGGGCTCCCACTGGTGTCTATCAGCCAGGCG 1527
451 ArgValAlaLeuGluAspGlyLysMetPheValThrProGlyAlaGlyAr 467
|||||
1528 AAGATTGTCTGGAGGCGCAGCTGTCATGTGCACGAAGGCTCAGCAGC 1577
467 gPheValProArgLysThrPheProAspPheValTyrLysArgIleLysA 484
|||||
1578 CTACATTCCTCCGAGGCTTCCCTGACTTGTGTACAAACGCATCAAGG 1627
484 laArgAsnArgLeuAlaGluIleHisGlyValProArgGlyLeuTyrAsp 500
|||||
1628 CAAGAGCAGGCTGGCTGAGCTGAGGGGGTCCCTGCTGGCTGTATGAT 1677
501 GlyProValHisGluValMetValProAlaLysProGlySerGlyAlaPr 517
|||||
1678 GGACCGGTATGCGAGGTGCTGTGAGCCCAAGACGGTCACTCCGCGCTC 1727
517 oAlaArgAlaSerCysProGlyLysIleSerValProValArgAsnL 534
|||||
1728 ATCAGCTAGACATCCCTTCGCAAGACGACGCGCCACCTGTTCGGAACC 1777
534 euHisGlnSerGlyPheSerLeuSerGlySerGlnAlaAspAspHisIle 550
|||||
1778 TGACCACTGCTGTTTACCTTGTCTGTGTCGTCAGATTGACGACCAACT 1827
551 AlaArgThrAlaGlnLysIleMetAlaProProGlyGlyArgSerAs 567
|||||
1828 CCCCAGCCAGCAGCAGCAGGATTTGTGGCGCCCTGTGTGGCGCGCA 1877
567 nIleThrSerLeu 571
|||||
1878 CATCAGGCGCTG 1890

```

seq_name: gb_pri10:HSU17279

seq_documentation_block: 1829 bp mRNA PRI 02-APR-1996

LOCUS HSU17279

DEFINITION Human collapsin response mediator protein hCRMP-2 mRNA, complete cds.

ACCESSION U17279

VERSION U17279.1 GI:1244399

KEYWORDS

SOURCE human.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 1829)

AUTHORS Goshima, Y., Nakamura, F., Strittmatter, P. and Strittmatter, S.M.

TITLE Collapsin-induced growth cone collapse mediated by an intracellular protein related to UNC-33

JOURNAL Nature 376 (6540), 509-514 (1995)

MEDLINE 95364923

REFERENCE 2 (bases 1 to 1829)

AUTHORS Strittmatter, S.M.

TITLE Direct Submission

JOURNAL Submitted (17-NOV-1994) Stephen M. Strittmatter, Neurology, Yale University School of Medicine, 333 Cedar Street, New Haven, CT 06520, USA

COMMENT On Apr 2, 1996 this sequence version replaced gi:882150.

FEATURES

source

1..1829

/organism="Homo sapiens"

/db_xref="taxon:9606"

72..1790

/note="collapsin response mediator protein"

/codon_start=1

/product="hCRMP-2"

/protein_id="AAA93202.1"

/db_xref="GI:1244400"

/translation="MSYQGKKNIPRTSDRLIKGKIVNDQSFYADIVMEDGLIKO IGENLIVPGVKTIEAHSRMVPGIDVHTRFQMPDQGMTSADDFQCTKAALAGTGT MIDHVVPEPCTSLAARDQWENADSKSCDYSLHVDI SEWHKGIQEMEAALYKDHG VNSFLVMAFKDRLQDLCIYEVLSVIRDIGAQAQVAENGDIIEEQQRILDIGIT GPESHVLSRPEVEAEVRAITIANQNCPLIITKNSKSAEVAIAQARKKGTVVYG EPITASLGTDSHYSKNKAFAAFVTPSPDPPTDFLNSLLSCDGLQVTGSAHC TFNTAQKAVGKDNFTLIPGNTGTEERMSVWDKAVTGMKMDNQFVAVTSTNAKVF NLYPRKGRIVAGSDADLVIPDPSVKTSIAKTHNSLEYNIPEGMCEKSGPLVTSQK KIVLEDGTLHVTEGSGRVIPRKPFDFVYIKRKARSRLAERGVPRGLVDGPGCEVSV TPKTVTPASSAKTSPAKQAPPPVRLHQSGFSLSGAQIDDDNIPTRTORIYVAPGGRA NITSIG"

BASE COUNT 426 a 509 c 511 g 383 t

ORIGIN

alignment_scores:

Quality: 2339.00 Length: 571

Ratio: 4.430 Gaps: 0

Percent Similarity: 92.469 Percent Identity: 75.482

alignment_block:

US-09-367-496-8 x HSU17279

Align seg 1/1 to: HSU17279 from: 1 to: 1829

1 MetSerPheGlnGlyLysSerIleProArgIleThrSerAspArgLe 17

72 ATGTCTTATCAGGGAAGAAAAATATTCACGCATCCAGCAGCATCGTCT 121

17 uleuileargGlyGlyArgIleValAsnAspAspGlnSerPheTyrAla 34

122 TCTGATCAAGAGGAGTAAATTTGTAATGATCAGCAGTCTGTCTATGCAG 171

34 spValHisValGluAspGlyLeuIleLysGlnIleGlyGluAsnLeuIle 50

172 ACATATACATGAGATGGTGTGATCAACCAANTAGGAGAAATCTGATT 221

```

51 valProGlyGlyIleHisThrIleAspAlaHisGlyLeuMetValLeuPr 67
|||||
222 GTCCAGGAGGAGTGAAGACCATCGAGGCCACTCCCGGATGGTATGCC 271
67 oGlyGlyValAspValHisThrArgLeuGlnMetProValLeuGlyMet 84
|||||
272 CGGAGGAATTCAGCTCCACACTCGTTCCAGATGCCCTGATCAGGAATGA 321
84 hrProAlaAspAspPheCysGlnGlyThrLysAlaAlaLeuAlaGlyGly 100
|||||
322 CGTCTGCTGATGATTTCTTCAAGGAACCAAGCGCCCTGGCTGGGGGA 371
101 ThrThrMetIleLeuAspHisValPheProAspPheGlyValSerLeuLe 117
|||||
372 ACCACTATGATCATGTACCACTGTTCTTCCAGCTGGGACCAACCTGCT 421
117 uAlaAlaTyrGluGlnTrpArgGluArgAlaAspSerAlaAlaCysCys 134
|||||
422 CGCTGCCTTTGACCAAGTGGAGGAATGGCCGACAGCAAGTCTGCTGTG 471
134 spYrSerLeuHisValAspIleThrArgTrpHisGluSerIleGlyGlu 150
|||||
472 ACTACTCTCTGATGTCGACATCAGCGAGTGGCATAAGGGCATCCAGGAG 521
151 GluLeuGluAlaLeuValLysGluLysGlyValAsnSerPheLeuValPh 167
|||||
522 GAGATGGAAGCGCTGTGNAAGATCAGCGGGTAATTCCTTCTCTGCTGTA 571
167 eMetAlaTyrLysAspArgCysGlnCysSerAspSerGlnMetTyrGluI 184
|||||
572 CATGGCTTTCAAGATCGCTCCAGCTAACGGAATTCAGATTTATGAAG 621
184 lePheSerIleIleArgAspLeuGlyAlaLeuAlaGlnValHisAlaGlu 200
|||||
622 TACTGAGTGTGATCCGGGATATTGGCCCATAGCCCAAGTCCACGCAGAA 671
201 AsnGlyAspIleValGluGluGlnLysArgLeuLeuGluGlyIle 217
|||||
672 ANTGGCCACATCATTCAGAGGACGACGAGGATCTCGGATCTGGGAT 721
217 eThrGlyProGluGlyHisValLeuSerHisProGluGluValGluAlaG 234
|||||
722 CACGGGCCGAGGAGCATGTGCTGAGCGGACCTGAGGAGGTCGAGGCGG 771
234 luAlaValTyrArgAlaValThrIleAlaLysGlnAlaAspCysProLeu 250
|||||
772 AAGCGGTGAATCGGCCATACCATCGCCCAACGACCAACTGCCCGCTG 821
251 TyrValThrLysValMetSerLysGlyAlaAlaAspAlaIleAlaGlnAl 267
|||||
822 TATATCACCAAGGTGATGAGCAAAAGCTCTGCTGAGGTCTATGCCAGGC 871
267 aLysArgGlyValValValPheGlyGluProIleThrAlaSerLeuG 284
|||||
872 ACGGAAGAAGGGAATGTTGTTGATGGGAGGCCCATCATCTGCCAGCTGG 921
284 lyThrAspGlySerHisTyrTrpSerLysAsnTrpAlaLysAlaAla 300
|||||
922 GAACGGAGGCTCCCATTTACTGGAGCAAGAAGTGGCCCAAGCTGCTGCC 971
301 PheValThrSerProValAsnProAspProThrThrAlaAspHisIle 317
|||||
972 TTTGTACCTCCCACTTGGAGGCTGATCCCAACCTCCAGACTTCT 1021
317 uThrCysLeuLeuSerSerGlyAspLeuGlnValThrGlySerAlaHisC 334
|||||
1022 CAACTCTCTGCTGCTGTGAGGACCTCAGGTACGGGCGAGTGCCTCAT 1071
334 ysThrPheThrThrAlaGlnLysAlaValGlyLysAspAspPheAlaLeu 350
|||||
1072 GCACGTTTAACATGCCAGAGGCTGTAGGAAGGACCACTTACACCTG 1121
351 IleProGluGlyThrAsnGlyIleGluGluArgMetSerMetValTrpG 367

```

```

|||||
1122 ATTCGGAGGCGACCAATGGCACTGAGGAGCGGATGTCGTCATCTGGA 1171
367 uLysCysValAlaSerGlyLysMetAspGluAsnGluPheValAlaValT 384
|||||
1172 CAAGGCTGTGTCACACTGGGAAGATGATGAGAACCAGTTTGTGGCTGTA 1221
384 hrSerThrAsnAlaAlaLysIlePheAsnPheTyrProArgLysGlyArg 400
|||||
1222 CCGACCAATGACCCAAAGTCTTCAACCTTTACCCCGGAAAGCGCGC 1271
401 ValAlaValGlySerAspAlaAspLeuValIleTrpAsnProLysAlaTh 417
|||||
1272 ATTGCTGTGGATCCGATGCGGACCTGCTCATCTGGACCCCGACAGCT 1321
417 rLysIleIleSerAlaLysThrHisAsnLeuAsnValGluTyrAsnIleP 434
|||||
1322 TAAACCATCTCTGCCAAGACACACACAGCTCTCTCGAGTACAAACATCT 1371
434 heGluGlyValGluCysArgGlyAlaProAlaValIleSerGlnGly 450
|||||
1372 TTGAAGGCTGTGAGTGGCGGCTCCCTCCACTGGTGTCTATCAGCCAGGG 1421
451 ArgValAlaLeuGluAspGlyLysMetPheValThrProGlyAlaGlyAr 467
|||||
1422 AAGATTGTCTGGAGGAGCGCACCTGCTCATGTCCAGGAAGGCTCTGGACG 1471
467 gPheValProArgLysThrPheProAspPheValTyrLysArgIleLysA 484
|||||
1472 CTACATTCCTCCGGAAGCCCTTCCCTGATTTGTTTACAGCGTATCAAGG 1521
484 laArgAsnArgLeuAlaGluIleHisGlyValProArgGlyLeuTyrAsp 500
|||||
1522 CAGGAGGAGGCTGGCTGAGCTGAGAGGGGTCTCTGCGCTGTATGAC 1571
501 GlyProValHisGluValMetValProAlaLysProGlySerGlyAlaPr 517
|||||
1572 GGNCCGTGTGTGAAGTGTCTGTGACGGCCCAAGACAGTCACTCCAGCCTC 1621
517 oAaArgAlaSerCysProGlyLysIleSerValProValArgAsnL 534
|||||
1622 CTCGCCCAAGAGCTCTCTGCCAAGCAGGAGGCCCTCCCTGTCGGAAAC 1671
534 euHisGlnSerGlyPheSerLeuSerGlySerGlnAlaAspAspHisIle 550
|||||
1672 TGCACCACTCTGATTCAGTTGCTGTGGTGTCTGATTCAGTATGATGACA 1721
551 AlaArgArgThrAlaGlnLysIleMetAlaProProGlyGlyArgSerAs 567
|||||
1722 CCGCCGCGCACCCAGCAGGTATCTGTGGCGCCCGCGGTGGCGTGCCAA 1771
567 nIleThrSerLeu 571
|||||
1772 CATCACCAAGCTG 1784

```

seq_name: gb_pr7:D78013

```

seq_documentation_block: 4459 bp mRNA PRI 05-FEB-1999
LOCUS D78013
DEFINITION Homo sapiens mRNA for dihydropyrimidinase related protein-2,
complete cds.
ACCESSION D78013
VERSION D78013.1 GI:1330239
KEYWORDS dihydropyrimidinase related protein-2; unc-33.
SOURCE Homo sapiens fetus brain cDNA to mRNA.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 4459)
AUTHORS Hamajima,N., Matsuda,K., Sakata,S., Tamaki,N., Sasaki,M. and
Nonaka,M.
TITLE Direct Submission
JOURNAL Submitted (20-OCT-1995) to the DDBJ/EMBL/GenBank databases. Naoki

```


1787 GAGATGGAGCGCTTGTGAAGATCATCGGGTAATTCCTTCCTCGTGTA 1836
167 eMetaTyrLysAspArgCysGlnCysSerAspSerGlnMetTyrGluI 184
1837 CATGGCTTCAAGATCCCTCCAGCTAAAGGATTCAGATTTATGAAG 1886
184 lePheSerIleileArgAspLeuGlyAlaLeuAlaGlnValHisAlaGlu 200
1887 TACTGAGTGTATCCGGATATTCGCCCATAGCCCAAGTCCAGCCAGAA 1936
201 AsnGlyAspIleValGluGluGlnLysArgLeuLeuGluLeuGlyI 217
1937 AATGGCAGCATCATTCAGAGGACACAGAGGATCTCGATCTGGGCAT 1986
217 eThrGlyProGluGlyHisValLeuSerHisProGluGluValGluAlaG 234
1987 CACGGGCGCGAGGAGATGTGTGAGCCGACCTGAGGAGGTGAGGCCG 2036
234 luAlaValTyrArgAlaValThrIleAlaLysGlnAlaAsnCysProLeu 250
2037 AGCCGTGAATCTGGCATCACCATCGCCCAACAGACCAACTGCCGCTG 2086
251 TyrValThrLysValMetSerLysGlyAlaAlaAspAlaIleAlaGlnAl 267
2087 TATATCACCAAGGTGATGACAAAGCTGTGCTGAGTCACTGCCAGGC 2136
267 aLysArgArgGlyValValValPheGlyGluProIleThrAlaSerLeuG 284
2137 ACGAAGAGGAAGTGTGTGTATGGCGAGCCCATCATCATCGAGCTGG 2186
284 lyThrAspGlySerHisTyrTrpSerLysAsnTrpAlaLysAlaAlaA 300
2187 GAACGGAGGCTCCCATTAAGGAGCAAGACTGGCCAGGCTGTGCTGCC 2236
301 PheValThrSerProValAsnProAspProThrThrAlaAspHisLe 317
2237 TTGTCACTCCCTCCCTGTAGCCCTGTATCCAACTCCAGACTTCT 2286
317 uThrCysLeuLeuSerSerGlyAspLeuGlnValThrGlySerAlaHisC 334
2287 CAACCTCTGTCTCTGTGGAGACTCCAGGTCACGGCAGGAGGCCAT 2336
334 ysThrPheThrThrAlaGlnLysAlaValGlyLysAspAsnPheAlaLeu 350
2337 GCAGCTTTAACTACCTGCCAAGAGGTGTAGGAAGGACAACTTACCCCTG 2386
351 IleProGluGlyThrAsnGlyIleGluArgMetSerMetValTrpG1 367
2387 ATTCGGAGGCGACCAATGGCACTGAGGAGCGGATGTCCGTCTCTGGGA 2436
367 uLysCysValAlaSerGlyLysMetAspGluAsnGluPheValAlaValT 384
2437 CAAGGCTGTGGTCACTGGGAAGATGGATGAGAACCATTTGTGGCTGTA 2486
384 hrSerThrAsnAlaAlaLysIlePheAsnPheTyrProArgLysGlyArg 400
2487 CCAGCCCAATACACCAAGTCTTCACTTACCCCGGAAAGGCCG 2536
401 ValAlaValGlySerAspAlaAspLeuValIleTrpAsnProLysAlaTh 417
2537 ATTGCTGTGGATCGATCGGACCTGTCTATCTGGGACCCCGCAGCGGT 2586
417 rLysIleIleSerAlaLysThrHisAsnLeuAsnValGluTyrAsnIleP 434
2587 TAAACCATCTCTGCAAGACACACACAGCTCTCTCGAGTACACATCT 2636
434 heGluGlyValGluCysArgGlyAlaProAlaValIleSerGlnGly 450
2637 TTGAAGCATGAGTGGCGGCTCCCTGCTGCTGCTGCTGCTGCTGCTG 2686
451 ArgValAlaLeuGluAspGlyLysMetPheValThrProGlyAlaGlyAr 467
2687 AAGATTGTCTGGAGGAGGACCCCTGCTGCTGCTGCTGCTGCTGCTG 2736

467 gPheValProArgLysThrPheProAspPheValTyrLysArgIleLysA 484
2737 CTACATCCCCGGAAGCCCTTCCCTGATTTGTTTACAAAGCGTATCAAG 2786
484 laArgAsnArgLeuAlaGluIleHisGlyValProArgGlyLeuTyrAsp 500
2787 CAAGGACGAGCTGGTGGAGTGGAGGGGTCTCTGGCTGCTGATGAC 2836
501 GlyProValHisGluValMetValProAlaLysProGlySerGlyAlaPr 517
2837 GGACCTGTGTGAAGTGTCTGTGACGCCCAAGACAGCTCACTCCAGCCTC 2886
517 oAlaArgAlaSerCysProGlyLysIleSerValProProValArgAsnL 534
2887 CTCGGCCCAAGACGCTCTCTGCCAAGCAGCAGGCCACCTCTCGGAACC 2936
534 euHisGlnSerGlyPheSerLeuSerGlySerGlnAlaAspAspHisIle 550
2937 TGCACCACTGTGGATTCAGTTGCTGTCTGTGCTCAGATTGATGACACAT 2986
551 AlaArgThrAlaGlnLysIleMetAlaProProGlyGlyArgSerAs 567
2987 CCCGCGCGCACCCACCCAGCGTATGCTGGCGCCGCCCGCTGCGCGCAA 3036
567 nIleThrSerLeu 571
3037 CATCACCGAGCTG 3049
seq_name: gb_patl:A87370
seq_documentation_block: 1817 bp DNA PAT 22-JAN-2000
LOCUS A87370 Sequence 1 from Patent WO9837192.
ACCESSION A87370
VERSION A87370.1 GI:6736126
KEYWORDS house mouse.
SOURCE Mus musculus
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 1817)
AUTHORS Byk,T. and Belin,M.
TITLE USE OF ULIP PROTEINS IN THE DIAGNOSIS AND THERAPY OF CANCER AND
PARANEOPLASTIC NEUROLOGICAL SYNDROMES
JOURNAL Patent: WO 9837192-A 1 27-AUG-1998;
BYK TAMARA (CH); BELIN MARIE FRANCOISE (FR)
FEATURES Location/Qualifiers
source 1..1817
/organism="Mus musculus"
/db_xref="taxon:10090"
BASE COUNT 426 a 506 c 511 g 374 t
ORIGIN

alignment_scores:
Quality: 2325.00 Length: 571
Ratio: 4.429 Gaps: 0
Percent Similarity: 91.944 Percent Identity: 75.306
alignment_block:
US-09-367-496-8 x A87370 ..
Align seg 1/1 to: A87370 from: 1 to: 1817
1 MetSerPheGlnGlyLysLysSerIleProArgIleThrSerAspArgLe 17
23 ATGCTTATCAGGGGAGCAAAATATTCCACCATCAGCAGGATCGTCT 72
17 uLeuIleArgGlyGlyArgIleValAsnAspGlnSerPheTyrAlaA 34
73 TCTGATCAAGGTGGCAAGATTTGATGATGATGATGATGATGATGATG 122


```

34 spValHisValGluAspGlyLeuIleGlnGlyLeuAsnLeuIle 50
123 ACATATACATGGAAGTGGTTGATCAAGCAATAGGAGAAACCTGATT 172
51 ValProGlyGlyIleHisThrIleAspAlaHisGlyLeuMetValLeuPr 67
173 GTACCAGAGGGGTGAAGACCATCGAAGCCACTCCAGAATGGTATTCC 222
67 oGlyGlyValAspValHisThrArgLeuGlnMetProValLeuGlyMet 84
223 CGGAGGAATTGACGTGTCATCTCGGTTCAGATGCTGACCAAGGAATGA 272
84 hrProAlaAspPheCysGlnGlyThrLysAlaAlaLeuAlaGlyGly 100
273 CATCCGCTGATGACATTCCTCCAGGGAACCAAGGCGGCTGCGCGGGA 322
101 ThrThrMetIleLeuAspHisValPheProAspThrGlyValSerLeu 117
323 ACCACCATGATCATTCAGCATGTTCTTCTGAGCCCGGACGAGCTATT 372
117 uAlaAlaTyrGluGlnTyrArgGluAlaAspSerAlaAlaCysCysA 134
373 GCGTCCCTTTGATCACTGAGGAGTGGCTGACAGCAAGTCCCTGCTGG 422
134 spTyrSerLeuHisValAspIleThrArgTrpHisGluSerIleLysGlu 150
423 ACTATTGCTGTCAGCTGACATCACTGAGTGGCAAGGCGCATCCAGGAG 472
151 GluLeuGluAlaLeuValLysGluLysGlyValAsnSerPheLeuValPh 167
473 GAGATGGAAGTCTGCTGGAAGGACCAAGCGGGTAAACTCTTCTCGTGT 522
167 eMetAlaTyrLysAspArgCysGlnCysSerAspSerGlnMetTyrGlu 184
523 CATGCTTTCAAGATGATTCACACTGACGAGGATTCACAGATCTATGAG 572
184 lePheSerIleIleArgAspLeuGlyAlaLeuAlaGlnValHisAlaGlu 200
573 TGCTGAGCGTATCGGATATCGGTGCAATAGTCAAGTCCACGACGAG 622
201 AsnGlyAspIleValGluGluGlnLysArgLeuLeuGluGlyIle 217
623 AATGGTGACATCATTTGCTGAGGACAGCAGAGGATCTTGGATCTGGCAT 672
217 eThrGlyProGlyHisValLeuSerHisProGluGluValGluAlaG 234
673 CACGGGCCCCAGGAGACAGTGTGAGCGGCGGAGAGGAGTCCAGGCTG 722
234 luAlaValTyrArgAlaValThrIleAlaLysGlnAlaAsnCysProLeu 250
723 AAGCTGTCAACCGTCCATCACTATTGCAACCAAGCAAGTCCGCTCTG 772
251 TyrValThrLysValMetSerLysGlyAlaAlaAspAlaIleAlaGlnAl 267
773 TATGTCACCAAGTATGCCCAAGAGTGGCGTGAAGTCATGCTCAGGC 822
267 aLysArgArgGlyValValValPheGlyGluProIleThrAlaSerLeu 284
823 ACGGAAGAGGGAAGTGTGTTGATGAGTGGAGCCCATCAGCGGACCCCTGG 872
284 lYThrAspGlySerHisTyrTrpSerLysAsnTrpAlaLysAlaAla 300
873 GGACTGATGGCTCTCATTTACTGGAGCAAGAACTGGGCAAGGCTCGCGCC 922
301 PheValThrSerProValAsnProAspProThrThrAlaAspHisLe 317
923 TTTGTACCTCCACCTTGGAGCCCGGACCAACCACTCCAGACTTTCT 972
317 uThrCysLeuLeuSerSerGlyAspLeuGlnValThrGlySerAlaHisC 334
973 CAACTGCTGTCTCTGAGACCTCCAAAGTCACTGGCAGTGGCCACT 1022
334 ysThrPheThrThrAlaGlnLysAlaValGlyLysAspAsnPheAlaLeu 350

```

```

1023 GCACCTTCAACACTGCCAGAGGCTGTGGGAAGCAACACTTCACTTG 1072
351 IleProGluGlyThrAsnGlyIleGluGluArgMetSerMetValTrpGl 367
1073 ATTCCGAGGGCACCACCAAGCTCTTCAACCTTTACCCCGGAAAGCTCG 1122
367 uLysCysValAlaSerGlyLysMetAspGluAsnGluPheValAlaVal 384
1123 TAAAGCTGTGGTCACTGGGAAGATGATGAGAATCAGTTTGTGGCTGTGA 1172
384 hrSerThrAsnAlaAlaLysIlePheAsnPheTyrProArgLysGlyArg 400
1173 CCAGCACCAACGACGAGCAAGTCTTCAACCTTTACCCCGGAAAGCTCG 1222
401 ValAlaValGlySerAspAlaAspLeuValIleTrpAsnProLysAlaTh 417
1223 ATCTCGTGGGATCTGATGCTGACTTGGTCACTCTGGGACCTGACAGTG 1272
417 rLysIleIleSerAlaLysThrHisAsnLeuAsnValGluTyrAsnIleP 434
1273 GAAGACCATCTCTGCCAAGACACACCAAGTCTTGTAGTACAACATCT 1322
434 heGluGlyValGluCysArgGlyAlaProAlaValValIleSerGlnGly 450
1323 TTGAAGGCATGGAGTCTCGCGGCTCCCACTGGTGTGTCATCAGCCAGG 1372
451 ArgValAlaLeuGluAspGlyLysMetPheValThrProGlyAlaGlyAr 467
1373 AAGATTGCTCTGGAGGACGACACTTTCATGCTCACTCAAGGCTCAGGAC 1422
467 gPheValProArgLysThrPheProAspPheValTyrLysArgIleLysA 484
1423 CTACATTCCTCCGGAAGCCCTTCCCTGACTTGTGTACAAACGATCAAAG 1472
484 laArgAsnArgLeuAlaGluIleHisGlyValProArgGlyLeuTyrAsp 500
1473 CAAGGAGCAGGCTGGCTGAGCTGAGAGGGTCCCTCGTGGCCTGTATGAC 1522
501 GlyProValHisGluValMetValProAlaLysProGlySerGlyAlaPr 517
1523 GGACCGGTATGCGAGGTGTCTGTGACGCCCAAGACGCTGACTCCAGCCT 1572
517 oAlaArgAlaSerCysProGlyLysIleSerValProProValArgAsnL 534
1573 ATCAGCTAAGACATCCCTGCCAAGCAGGACGACCCCTGTTTCGGAACC 1622
534 euHisGlnSerGlyPheSerLeuSerGlySerGlnAlaAspAspHisIle 550
1623 TGCACCACTCTGGATTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1672
551 AlaArgArgThrAlaGlnLysIleMetAlaProProGlyGlyArgSerAs 567
1673 CCCTCGCGGACCAACGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1722
567 nIleThrSerLeu 571
1723 CATCACCAAGCTG 1735

```


* Tue Jul 31 13:08:21 2001

OM of: US-09-367-496-8 to: EST:* out_format : pfs

Date: Jul 30, 2001 12:07 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODE=frame-p2n.model -DEV=xlpl
-O=/cgn2.1/USPTO_spool/US09367496/runat_30072001_081220_27362/app_query.fasta_1.637
-DB=EST -FWT=fastap -SUFFIX=std.rst -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPCT=0.000 -LOOPEXT=0.000 -QGAPOP=4.500
-QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -DELOP=6.000
-FGAPOP=7.000 -XGAPOP=10.000 -XGAPEXT=0.500 -DELOP=6.000
-DELEXT=7.000 -START=1 -MATRIX=blossum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THRSFMT=pct -THR_MAX=100 -THR_MIN=0
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0
-MAXLEN=200000000 -USER=US09367496.@CGN1_1_3359 -NCPU=6
-ICPU=3 -LONGLOG -NO_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-367-496-8

Query length: 572

Database: EST:*

Database sequences: 10228115

Database length: 431459454

Search time (sec): 1362.640000

score_list:

Sequence	Strd	Orig	zScore	Escore	Len	Documentation
gb_hlc:AK004822	+	1518.50	2772.79	2.5e-145	2109	AK004822 Mus musculus adult ma
gb_hlc:AK004899	+	1515.00	2764.41	7.3e-145	2471	AK004899 Mus musculus adult ma
gb_est29:AL535253	+	1218.00	2226.32	6.9e-115	989	AL535253 LTI_FL013_FBI
gb_est29:AL538995	+	1103.00	2014.50	4.3e-103	925	AL538995 LTI_FL013_FBI
gb_est29:AL538995	+	1101.50	2011.86	6.1e-103	925	BF530817 602072175F1 NCI-CGAP
gb_est29:AL538995	+	1091.00	1992.38	7.4e-102	920	AL518319 LTI_FL013_FBI
gb_est29:AL538995	+	1068.00	1951.18	1.3e-99	778	AL538995 LTI_FL013_FBI
gb_est29:AL538995	+	1046.00	1913.13	1.9e-97	663	AA778518 af85905.sl Soares Test
gb_est29:AL538995	+	1029.00	1882.67	9.5e-96	612	AA778518 af85905.sl Soares Test
gb_est29:AL538995	+	1025.00	1875.67	2.3e-95	592	AA778518 af85905.sl Soares Test
gb_est29:AL538995	+	1025.00	1868.77	5.7e-95	1147	AL535160 AL535160 LTI_FL013_FBI
gb_est29:AL538995	+	1015.00	1849.53	6.7e-94	1117	AL535160 AL535160 LTI_FL013_FBI
gb_est29:AL538995	+	995.00	1818.23	3.7e-92	697	AL535160 AL535160 LTI_FL013_FBI
gb_est29:AL538995	+	992.50	1812.36	7.9e-92	773	AL535160 AL535160 LTI_FL013_FBI
gb_est29:AL538995	+	985.00	1798.32	4.8e-91	784	AL535160 AL535160 LTI_FL013_FBI
gb_est29:AL538995	+	984.50	1799.19	4.3e-91	676	AL535160 AL535160 LTI_FL013_FBI
gb_est29:AL538995	+	959.00	1752.27	2.2e-90	971	AL535160 AL535160 LTI_FL013_FBI
gb_est29:AL538995	+	957.00	1748.45	2.9e-88	663	AL535160 AL535160 LTI_FL013_FBI
gb_est29:AL538995	+	946.00	1727.41	4.2e-87	710	AL535160 AL535160 LTI_FL013_FBI
gb_est29:AL538995	+	930.50	1695.63	2.5e-85	920	AL535160 AL535160 LTI_FL013_FBI
gb_est29:AL538995	+	928.00	1691.48	4.3e-85	885	AL535160 AL535160 LTI_FL013_FBI
gb_est29:AL538995	+	925.50	1686.57	8.0e-85	906	AL535160 AL535160 LTI_FL013_FBI
gb_est29:AL538995	+	924.50	1685.85	8.8e-85	825	AL535160 AL535160 LTI_FL013_FBI
gb_est29:AL538995	+	924.00	1686.24	8.3e-85	740	AL535160 AL535160 LTI_FL013_FBI
gb_est29:AL538995	+	920.00	1679.99	1.9e-84	673	AL535160 AL535160 LTI_FL013_FBI
gb_est29:AL538995	+	910.00	1650.49	8.2e-83	1678	AL535160 AL535160 LTI_FL013_FBI
gb_est29:AL538995	+	898.50	1638.20	4.0e-82	797	AL535160 AL535160 LTI_FL013_FBI
gb_est29:AL538995	+	893.50	1627.54	1.5e-81	896	AL535160 AL535160 LTI_FL013_FBI
gb_est29:AL538995	+	890.00	1623.26	2.7e-81	747	AL535160 AL535160 LTI_FL013_FBI
gb_est29:AL538995	+	885.00	1611.91	1.2e-80	890	AL535160 AL535160 LTI_FL013_FBI
gb_est29:AL538995	+	881.00	1604.03	3.2e-80	926	AL535160 AL535160 LTI_FL013_FBI
gb_est29:AL538995	+	878.00	1598.09	4.6e-80	870	AL535160 AL535160 LTI_FL013_FBI
gb_est29:AL538995	+	879.00	1590.04	6.8e-80	961	AL535160 AL535160 LTI_FL013_FBI
gb_est29:AL538995	+	877.00	1585.33	9.7e-80	956	AL535160 AL535160 LTI_FL013_FBI
gb_est29:AL538995	+	876.50	1581.80	5.5e-79	682	AL535160 AL535160 LTI_FL013_FBI
gb_est29:AL538995	+	867.00	1581.80	5.5e-79	682	AL535160 AL535160 LTI_FL013_FBI
gb_est29:AL538995	+	866.00	1577.47	6.6e-79	840	AL535160 AL535160 LTI_FL013_FBI
gb_est29:AL538995	+	866.00	1577.47	6.6e-79	840	AL535160 AL535160 LTI_FL013_FBI
gb_est29:AL538995	+	866.00	1567.06	3.6e-78	857	AL535160 AL535160 LTI_FL013_FBI
gb_est29:AL538995	+	852.00	1549.55	3.4e-77	995	AL535160 AL535160 LTI_FL013_FBI
gb_est29:AL538995	+	848.00	1547.41	4.5e-77	643	AL535160 AL535160 LTI_FL013_FBI

gb_est91:BF685809 + 846.50 1541.48 9.6e-77 836 + BF685809 602140492F1 NIH_MGC
gb_est29:AL528054 + 846.00 1539.46 1.2e-76 915 + AL528054 AL528054 LTI_FL013_FBI
gb_est29:AL528054 + 844.00 1538.34 1.4e-76 739 + BE952303 UI-M-CD0-ayl-b-04-0
gb_est29:AL535518 + 844.00 1535.98 2.0e-76 899 + AL535518 AL535518 LTI_FL013_FBI

seq_name: gb_hlc:AK004822

seq_documentation_block: 2109 bp mRNA HTC 08-FEB-2001
LOCUS AK004822
Mus musculus adult male lung cDNA, RIKEN full-length enriched
library, clone:1200017110, full insert sequence.

ACCESSION AK004822.1 GI:12836294

VERSION CAP trapper.

KEYWORDS Mus musculus (strain:C57BL/6J) adult male lung cDNA to mRNA,

SOURCE muscle_lib:RIKEN full-length enriched mouse cDNA library

clone:1200017110.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

1 (sites)

Carninci,P. and Hayashizaki,Y.

TITLE High-efficiency full-length cDNA cloning

JOURNAL Methods Enzymol. 303, 19-44 (1999)

REFERENCE 2 (sites)

Carninci,P. and Hayashizaki,Y.

TITLE Normalization and subtraction of cap-trapper-selected cDNAs to

JOURNAL prepare full-length cDNA libraries for rapid discovery of new genes

TITLE Genom Res. 10 (10), 1617-1630 (2000)

JOURNAL 20499374

MEDLINE 3 (sites)

Shibata,K., Itoh,M., Aizawa,K., Nagao,K., Sasaki,N., Carninci,P.,

Konno,H., Akiyama,J., Nishi,K., Kitsuai,T., Tashiro,H., Itoh,M.,

Kikuchi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T.,

Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T.,

Kashiwagi,K., Fujiiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,

Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T.,

Matsui,K., Okazaki,Y., Muramatsu,M., Inoue,Y. and Hayashizaki,Y.

RIKEN integrated sequence analysis (RISA) system -384-format

sequencing pipeline with 384 multicapillary sequencer

Genom Res. 10 (11), 1757-1771 (2000)

20530913

MEDLINE 4 (sites)

The RIKEN Genome Exploration Research Group Phase II Team and

FANTOM Consortium.

TITLE Functional annotation of a full-length mouse cDNA collection

JOURNAL Nature 409, 685-690 (2001)

REFERENCE 5 (bases 1 to 2109)

Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,

Arakawa,T., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M.,

Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F.,

Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kato,H., Kawai,J.,

Kojima,Y., Konno,H., Kouda,K., Koyama,K., Kurahara,C., Matsuyama,T.,

Miyazaki,A., Nishi,K., Nomura,K., Numasaki,R., Ohno,M., Okazaki,Y.,

Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sano,H.,

Sasaki,D., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T.,

Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,

Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A.,

Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.

Direct Submission

TITLE Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of

Physical and Chemical Research (RIKEN), Laboratory for Genome

Exploration Research Group, RIKEN Genomic Sciences Center (GSC),

RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,

Kanagawa 230-0045, Japan (E-mail:genome-res@gs.riken.go.jp,

URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,

Fax:81-45-503-9216)

Please visit our web site (http://genome.gsc.riken.go.jp/) for

further details

cDNA library was prepared and sequenced in Mouse Genome

Encyclopedia Project of Genome Exploration Research Group in Riken

Genomic Sciences Center and Genome Science Laboratory in RIKEN.

COMMENT


```

/translation="MAPQRLIRGRIVNDDFSQVADLVEDGVVVRALGRDLLPPED
ASRGLRIIDAAAGKLVLPGLIDHTHMQPFPMGSSQVDDFYQCTKAALAGTMTIIDFA
IPQKSSLIIEAFETWRNADPKVCCDYSILHVAVTWMSDKVKEEMKTLARDGVNSFKM
FMAYKGLYVWDOEOLYAAFSOCKEAGIAVQHAENGDLIAEGAKTLALGITGPGHE
LCRPVABEATLRAITASAVNCPPLYVHVMSKSAARVADARRAGNVVYGEPTAAG
LGTGROYSEWSESHAAHVHVPPLDPPLTPGLMDLLANGDLTTTGSNDNCTFNCTQ
KALGKDDTDKIPNGVNGVEDRMSVLEWEGHSGRMDENRFVAVTSTNAKIFNLYPKK
GRIAVGSDIIVNDPEATRISAKTHQAVNFNIFEGMVCHGVPLVTISRGRVVEA
GVFNVTAGHGKFIPIRQPAEYIYKRIKORDQCTPVPVKRAPYKGEVTLKARETKED
DTAGTRMQGHS"
polyA_signal      /note="putative"
2448..2453
polyA_site        /note="putative"
2471
BASE COUNT      682 a 583 c 636 g 570 t
ORIGIN

```

```

alignment_scores:
  Quality: 1515.00      Length: 492
  Ratio: 3.788          Gaps: 1
  Percent Similarity: 81.301  Percent Identity: 58.537
alignment_block:
US-09-367-496-8 x AK004899 ..
Align seg 1/1 to: AK004899 from: 1 to: 2471
16 ArgLeuLeuLeuLeuArgGlyGlyValArgIleValAsnAspGlnSerPheThy 32
124 CGACTTCATCCCGGGGGTGCATTCATCAATGACGACTTCTCACAGGT 173
32 rAlaAspValHisValGluAspGlyLeuIleValGlyGlnIleGlyGluAsnL 49
174 GCCGATGCTAGTGGAGGACGGCGTGGCGGGCTGGGCGGCGTGGAGCGGGACC 223
49 euIleValPro.....GlyGlyIleHisThrIleAlaHis 61
224 TGCTGCCTCTCAGGACGACATCTCGGGGGTGCAGAACTCTAGATGCACGG 273
62 GlyLeuMetValLeuProGlyGlyValAspValHisThrArgLeuGlnMe 78
274 GGCAAGCTGCTCCGCGGGGGCATCGACGACGACGACGACATGCAAT 323
78 tProValLeuGlyMetThrProAlaAspPheCysGlnGlyThrLysA 95
324 CCCGTTTCCGCTCGCAGTCAGTGGAGGACTTCTACCGAGGACCAAG 373
95 laAlaLeuAlaGlyGlyThrMetIleLeuAspHisValPheProAsp 111
374 CTGCTTTGGCAGGAGGACCAACCATGATCATGATTTGGCATTTCTCTCAG 423
112 ThrGlyValSerLeuLeuAlaAlaValArgGlnThrArgGluArgAlaAs 128
424 AAAGGACAGCTCCCTCATCGAGGCTTTGACACTTGGCGCACTGGGCGAGA 473
128 pSerAlaAlaCysCysAspTyrSerLeuHisValAlaSpilleThrArgTrp 145
474 CCGCAAGTCTGCTGACTATACCTGTCATGCTGCGAGTGGCAGTGCAGTGG 523
145 isGluSerIleLeuGluLeuGluAlaLeuValLysGluLysGlyVal 161
524 GCGACAAGGTAAAGAAAGAAATGAAACCTTGCCCGAGATAAAGCGGT 573
162 AsnSerPheLeuValPheMetAlaTyrLysAspArgCysGlnCysSerAs 178
574 AACTCTTTCAAGATGTTTATGGCTTACAAAGGCTCTGTACATGTGTGCAAG 623
178 pSerGlnMetTyrGluIlePheSerIleIleArgAspLeuGlyAlaLeuA 195
624 CGAGCAGCTGTATGGCGCTTCTCAGTGGCAAGGAGATAGAGGAGATG 673
195 laGlnValHisAlaGluAsnGlyAspIleValGluGluGlnLysArg 211

```

```

|||||
674 CTCAGGTGCAGCGCGAGATGGAGATTGTATGTCAGAGGGAGCAAGAAG 723
212 LeuLeuGluLeuGlyIleThrGlyProGluGlyHisValLeuSerHisPr 228
724 ACCTGGCAGTGGGATTAACAGGCCCTTGGAGGGCAGCAGCTGTGCCGTCC 773
228 oGluGluValGluAlaGluAlaValTyrArgAlaValThrIleAlaLysG 245
774 AGAAGCAGTGGAGGAGAGCCACCCTGAGAGCCATCACCATTAGCCAGCG 823
245 lnAlaAsnCysProLeuTyrValThrLysValMetSerLysGlyAlaAla 261
824 CTGTGAACCTGCCCCCTTATACGTGCTGATGTGATGAGCAAAATCTGACGC 873
262 AspAlaIleAlaGlnAlaLysArgArgGlyValValValPheGlyGluPr 278
874 AAGTGGTAGCGATGCGAGGAGAGCAAAATGTGCTTATGGAGAAC 923
278 orIleThrAlaSerLeuGlyThrAspGlySerHisTyrTrpSerLysAsn 295
924 AATTCAGCGGGTCTCGGCACCTGATGGCAGACAGTACTGGAGTGAAGAAT 973
295 rpAlaLysAlaAlaAlaPheValThrSerProProValAsnProAspPro 311
974 GGAGCCATGCGAGCCACCACCATGTCATGGTGGTCCGCCACCTGAGACCCGATCCT 1023
312 ThrThrAlaAspHisLeuThrCysLeuLeuSerSerGlyAspLeuGlnVa 328
1024 TTAACACCTGGCTTTCTTATGATCTGTGGCTAATGGGATTTGACCCAC 1073
328 lThrGlySerAlaHisCysThrPheThrThrAlaGlnLysAlaValGlyL 345
1074 AACAGGAGTGCACACTCTTTTAAACCTTGCACAAAGCTCTAGGGA 1123
345 ysAspAsnPheAlaLeuLeuProGluGlyThrAsnGlyIleGluGluArg 361
1124 AGGATGACTTCACTAAGATTCCCAATGGGTGAATGTTGTCGAGGACAGG 1173
362 MetSerMetValTrpGluLysCysValAlaSerGlyLysMetAspGluAs 378
1174 ATGTCGGTATATGGAAAGGGCGTCACAGTGTGTAATAATGGTGAAGA 1223
378 nGluPheValAlaValThrSerThrAsnAlaAlaLysIlePheAsnPhe 395
1224 CAGTTTGTGGCAGTTACCAGCACAACGACGCAAAATCTTTAATCTTT 1273
395 yrProArgLysGlyArgValAlaValGlySerAspAlaAspLeuValIle 411
1274 ATCCGAAAAGGAAGAAATAGCTGTAGGCTCAGATGCTGACATTGTGATC 1323
412 TrpAsnProLysAlaThrLysIleIleSerAlaLysThrHisAsnLeuAs 428
1324 TGGACCCAGACAACAAGAGGATCTCAGCCAAAATCATCATCAGGC 1373
428 nValGluTyrAsnIlePheGluGlyValGluCysArgGlyAlaProAlaV 445
1374 CGTTAATCTCAACATCTTCGAGGGCATGTTTGTACGGGTGCCCTGG 1423
445 alValIleSerGlnGlyArgValAlaLeuLeuGluAspGlyLysMetPheVal 461
1424 TGACTATTTCAAGAGCAGAGTGGTGTATGAGCAGGTGTTTTCACAGCTC 1473
462 ThrProGlyAlaGlyArgPheValProArgLysThrPheProAspPheVa 478
1474 ACAGCAGGACATGGGAAGTTTATTCGCCGCCAACCTTTTGGCCGAATACAT 1523
478 lTyrLysArgIleLysAlaArgAsnArgLeuAlaGluIleHisGlyValP 495
1524 TTACAAACCAATCAAGCAGGACGACGACCTGCACACCTGTGCCGTGA 1573
495 roArgGlyLeuTyrAspGlyProVal 503
|||||

```

203	AGCTGACACCAATCTGCTGTATTACTCCCTCCAGTGCACATCACAA	252
143	rgtPhHisGluSerIleLySglucluleuGluAlaLeuVallyLySgluLyS	159
253	GCTGGTACGATGGCGTTCGGGAGAGAGTGGAGGTGCTGGTGCAGGACAA	302
160	GlyValAsnSerPheLeuValPheMetAlaTyrLySAspArgCysGlnCy	176
303	GGCGTCAATTCCTCCAGTCTACATGGCCTATAAGGATGTCTACCAAA	352
176	sSerAspSerGlnMetTyrGluIlePheSerIleIleArgAspLeuGlyA	193
353	GTCCGACAGCCAGCTCTATGAAGCCTTTTACCTCTCTTAAGGCGCTGGAG	402
193	IaLeuAlaGlnValHisAlaGluAsnGlyAspIleValGluGluGln	209
403	CTGTGATCTTGTCTCATGCAGAAATGAGATTTTATAGCTCAGGAACAA	452
210	LySArgLeuLeuGluLeuGlyIleThrGlyProGluGlyHisValLeuSe	226
453	AAGCGGATCTCGAGATGGGCATCAGGGTCCGAGGGCCATGCCCTGAG	502
226	rHisProGluGluValGluAlaGluAlaValTyrArgAlaValThrIleA	243
503	CAGCCTCTGAAGAGCTGGAGCCGAGCGCTGTCCGGGCGCATCACCATTG	552
243	IaLySAlaAsnCysProLeuTyrValThrLySValMetSerLySgLy	259
553	CGGCGCGGATCAACTGCCCTGTGTATCACCAGGTCATGACCAAGAT	602
260	AlaAlaAspAlaIleAlaGlnAlaLySArgArgGlyValValPheG1	276
603	GCAGCCGACATCATCGCTCTGGCCAGGAGAAAGGGGCCCTAGTTTTGG	652
276	yGluProIleThrAlaSerLeuGlyThrAspGlySerHisTyrTipSerL	293
653	AGAGCCCATTTGCCGCCAGCCTGGGACCGATGGCAGCCCATCTACTGGACCA	702
293	yAsnTTPAlaLySAlaAlaAlaPheValThrSerProProValAsnPro	309
703	AGNACTGGGCCAAGCTCGGCGTTCGTGACTTCCTCCCTCCCTGAGCGCG	752
310	AspProThrThrAlaAspHisLeuThrCysLeuLeuSerSerGlyAspLe	326
753	GACCTCACACGCCGCGGACTACTTGACCTCCCTACTGGCTGTGGGCACTT	802
326	uclnValThrGlySerAlaHisCysThrPheThrThrAlaGlnLySAlaV	343
803	GCAGGTACAGGCACGGCCACTGTCCTACAGCACTGCCCAAGAGCGGG	852
343	alGlyLyAspAsnPheAlaLeuIleProGluGlyThrAsnGlyIleGlu	359
853	TGGGCAAGGACACACTTACCTCGATCCCGGAGGTGTCAACGGGATAGAG	902
360	GluArgMetSerMetValTTPGluLySAspValAlaSerGlyLyS	374
903	GACCGGATCAGGTCGCTCTGGGACAGGCGGTGGTACTGGCAAA	947

seq_name: gb_est29:AL538995
seq_documentation_block:
LOCUS AL538995 925 bp mRNA EST
DEFINITION AL538995 LTI_FL013.Fbrn1 Homo sapiens cDNA clone CS1 prime, mRNA sequence.
ACCESSION AL538995
VERSION AL538995.1 GI:12867812
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
REFERENCE 1 (bases 1 to 925)

```

1574 AGCGGGCCCTTACAGGGGAGAGTTC 1599
seq_name: gb_est29:AL535253

seq_documentation_block: 989 bp mRNA EST 13-FEB-2001
AL535253 LOCUS
AL535253 LTI_FL013_FBRn1 Homo sapiens cDNA clone CS0DF008YF05 5
prime, mRNA sequence.
AL535253
AL535253.1 GI:12798746
EST.
human.
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Li.W.B., Gruber.C., Jessee.J. and Polayes.D.
Full-length cDNA libraries and normalization
Unpublished (2001)
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
Location/Qualifiers
1..989
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DF008YF05"
/clone_lib="LTI_FL013_FBRn1"
/dev_stage="pooled tissue from post conception fetuses (20
week, 24 week and 26 week)"
/lab_host="DH10B"
/notes="Organ: Fetal brain; Vector: pCMVSPORT 6; 1st strand
cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-stranded cDNA was digested with Not I
and cloned into the Not I and Eco RV sites of the
PCMVSPORT 6 vector. Library was constructed by Life
Technologies. Contact : Feng Liang Life Technologies, a
division of Invitrogen 9800 Medical Center Drive Rockville
, Maryland 20850, USA Fax : (1) 301 610 8371 Email :
fliang@lifetech.com URL :
http://fulllength.invitrogen.com"
BASE COUNT 224 a 276 c 296 g 192 t 1 others
ORIGIN

alignment_scores:
Quality: 1218.00 Length: 315
Ratio: 4.289 Gaps: 0
Percent Similarity: 90.159 Percent Identity: 69.841

alignment_block:
US-09-367-496-x AL535253 ..
Align seg 1/1 to: AL535253 from: 1 to: 989

60 AlAHISgLyLeuMetValLeuProGlyGlyValAspValHisThrArgLe 76
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
3 GCCAACGGGGCGATGGTTATTCCGGAGGATTGATGTCACACGTCACCT 52
76 uGlnMetProValLeuGlyMetThrProAlaAspPheCysGlnGlyT 93
||||| ||| ||||||||| ||||||||| ||||||||| |||||||
53 GCAGAAAGCCCTCCACAGGGGATGATCGCGGTGATGACTTCTCCAAAGGA 102
93 hrLySAlaAlaLeuAlaGlyGlyThrMetIleLeuAspHisValPhe 109
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
103 CCAGGGCGGCACTGGTGGCGGGACACGATGATCATTCACCATGTGTT 152
110 ProAspThrGlyValSerLeuLeuAlaAlaTrpGluGlnTrpArgGluAr 126
|||||: ||| ||||||||| ||||||||| ||||||||| |||||||
153 CCTGACCTGGGTCCGACCTACTGACCTCTTCGAGAAAGTGGCAGGAAGC 202
126 qAlaAspSerAlaAlaCysCysAspTyrSerLeuHisValAspIleThr 143

```



```

loss_tissue_type="anaplastic oligodendroglioma with lp/19q
loss_tab_host="DH10B (T1 phage-resistant)"
/note_1="Organ: Brain; Vector: PCMV-SPORT6; Site_1: NotI;
/note_2="Site: Sali; Cloned unidirectionally. primer: Oligo dt.
Average insert size 2.3 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT      197 a      271 c      311 g      136 t
ORIGIN

```

```
alignment_scores:
  Quality: 1101.50
  Gaps: 4
  Length: 269
  Ratio: 4.337
```

Percent Similarity: 94.424 Percent Identity: 91.022

```
alignment_block:
US-09-367-496-8 x BF530817 ..
Align seg 1/1 to: BF530817 from: 1 to: 915
```

50 IleValProGlyGlyIleHisThrIleAspAlaHisGlyLeuMetVal 66
|||||
3 ATCGTCCTGGGGGCATCAAGACCATTTGACCCACGGCTG.ATGGTCCT 51

66 uProGlyGlyValAspValHisThrArgLeuGlnMetProValLeuClyM 83

52 TCCTGGTGGCGGTTACAGCCTGTCTGTCATGAGGCAAGGCGGCCTACGAGA 151
83 etThrProAlaAspPheCysGlnGlyThrLysAlaAlaLeuAlaCly 99
103 TGGTCCCCGCCAGCATTTCTTCAGGCACCAGSCAGCGCTACGAGA 151

107 TGACACC GGCCGAGCGCTTCGGTGGGT
100 GlyThrThrMetIleLeuAspHisValPheProAspThrGlyValSerLeu 116
|||||

[illegible]

116 uueuAALatYtGluuuniripLys 1111
202 GCTGGCGGCCTACGAGCAGTGGCGGAGCCGGGGACAGCGCGGCG. TGCT 250
133 ysAspTyrSerLeuHisValAspileThrArgTrpPHisGluSerIleLys 149

251 GGCACCTACTCCCTGTCACGCTGGACATCACCCGATGGCATGAGAGCATCAAG 300

150 GluCluLeuGluAlaLeuValLysGlyGlyValAsnSerPheLeuVa 166

301 GAGGACCTGGAGGCCCTGGTCAAGGAGAAGGCTGTGAACCTCTCTCGGT 350

166 lPheMetAlaTyLysAspArgCysGlnCysSerAspSerGlnMetTyrg 183

351 CTTTCATGGCATAACAAGGACCGGTGGCTGCAGTGCAGGACGACGCAGATGTACG 400

401 AGATCTTCAGCATCATCCGGGACCTGGGGGCCTTGGCCCCAGGTGCACGCT 450
|||||
200 CCGAACCAGVAsnTLeValGcuglUGluGlnLysArgLeuLeuGluLeuCl 216

451 GAGAACGGGGGACATCGTGTGGAGGAGAGGCACGATCTGGGGGGGGGG
216 yIeThrClyProGluClyHisValLeuSerHisProGluGluValGluA 233

233 laGluAlaValTyrArgAlaValThrIleAlaLysGlnAlaasnCysPro 249
|||||
551 CTGAGCGGGTGTACCGAGCTGTACCATCGCCAAAGCAGC.AACTGCCCG 599

250 LeuTyrValThrLysValMetSerLysGlyAlaAlaAspAlaIleAlaGI 266
|||||

600 CTGTACGTCACCACAAGGTGGTGAAGCAGCGCCCCCCCCTATCCGCC

```

3  AAGAGTGCAGCCGACATCATCTGCTGGCCAGGAAAGGCGCCCTAGT 52
274 lPheGlyGluProIleThrAlaSerLeuGlyThrAspGlySerHisTyrT 291
53  TTTTGGAGAGCCCATTTGCCCGCAGCTGGGACCGATGCCACCATTTACT 102
291 rPserLysAsnTrpAlaLysAlaAlaIlePheValThrSerProVal 307
103 GGAGCAAGAACTGGCCAAAGCTGGGGTTCGTGACTTCCCTCCCTG 152
308 AsnProAspProThrThrAlaAspHisLeuThrCysLeuLeuSerG 324
153 AGCCCGGACCTTACCACGCGGCTACTACTGACCTCCCTACTGGCTTGG 202
324 yAspLeuGlnValThrGlySerAlaHisCysThrPheThrThrAlaGlnL 341
203 GGACTTGCAGTGCAGAGCGGCGGCTGTCCTACAGCACTGCCGAGA 252
341 yAlaValGlyLysAspAsnPheAlaLeuIleProGluGlyThrAsnGly 357
253 AGCGGTGGGCAAGGACACTTACCTGTATCCCGGAGGTGTCAACGGG 302
358 IleGluGluArgMetSerMetValTrpGluLysCysValAlaSerGly 374
303 ATAGAGGAGCGGATGACGGTCTGCTGGCAAGCGGTGCTACTGGCAA 352
374 sMetAspGluAsnGluPheValAlaValThrSerThrAsnAlaLysI 391
353 AATGGATGAGAACCACTTGTGCTGTACCAAGCAATGACGCAAGA 402
391 lPheAsnPheThrProArgLysGlyArgValAlaValGlySerAspAla 407
403 TCATTAACTGTACCAAGCAAGGCGGATGCTGGTGGCTCGGATGCC 452
408 AspLeuValIleTrpAsnProLysAlaThrLysIleIleSerAlaLys 424
453 GACGTGCTCATCTGGACCCGACCAAGTTGAAGACCATAACAGCCAA 502
424 rHisAsnLeuAsnValGluTyrAsnIlePheGluGlyValGluCysArg 441
503 TCACAAGTGGCGGTGGAGTACAACTTTCGAGGATGAGTGCCAGC 552
441 lYAlaProAlaValIleSerGlnGlyArgValAlaLeuGluAspGly 457
553 GCTCCCACTAGTGTATCATCAGCAGGCGGATGCTTTTGAAGACGA 602
458 LysMetPheValThrProGlyAlaGlyArgPheValProArgLysThr 474
603 AACATCAACGTCAACAAGGCGATGGCGCTTCATTCGCGGAGGCGTT 652
474 eProAspPheValThrLysArgIleLysAlaArgAsnArgLeuAlaGlu 491
653 CCGGAGCACTGTACCAAGCGGCTCAAAATCAGGAATAAGTTTGGAT 702
491 lHisGlyValProArgGlyLeuTyrAspGlyProValHisGluValMet 507
703 TGCAAGGGTTCACAGGCGATGTATGACGGTCTGTGTACAGGATACA 752
508 ValProAlaLysProGlySerGlyAlaProAlaArgAlaSerCysPro 524
753 GCTACACCAATATCACTCCCGCTCTTCCAGCAATCTTCGCTTC 802
524 yLysIleSerValProProValArgAsnLeuHisGlnSerGlyPheSer 541
803 TAAACACAGCCCGCCACCATCAGAAACCTCCACCACTTCACCTTCAG 852
541 eSerGlySerGlnAlaAspHisIleAlaArgArgThrAlaGlnLys 557
853 TATCAGGTCCCGCAGATAGATCAACAATCCAGCGCACGCGG.CACGG 901
558 lLeMetAlaProPro 562
902 ATCTGGGCGCCCGC 916

```

seq_name: gb_est29.AL533854

seq_documentation_block:

LOCUS AL533854 778 bp mRNA EST 13-FEB-2001
 DEFINITION AL533854 LTI_FL013_FBrnl Homo sapiens cDNA clone CS0DF003YA05 5
 prime, mRNA sequence.
 ACCESSION AL533854
 VERSION AL533854.1 GI:12797347
 KEYWORDS EST
 SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 778)
 AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
 TITLE Full-length cDNA libraries and normalization
 JOURNAL Unpublished (2001)
 COMMENT Contact: Genoscope

Genoscope - Centre National de Sequencage

BP 191 91006 EVRY cedex - France

Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES

source

1..778
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="CS0DF003YA05"
 /clone_lib="LTI_FL013_FBrnl"
 /dev_stage="pooled tissue from post conception fetuses (20
 week_24 week and 26 week)"
 /lab_host="DH10B"
 /note="Organ: Fetal brain; Vector: pCMVSPORT 6; 1st strand
 cDNA was primed with a NotI-oligo(dT) primer. Five prime
 end enriched, double-stranded cDNA was digested with Not I
 and cloned into the Not I and Eco RV sites of the
 pCMVSPORT 6 vector. Library was constructed by Life
 Technologies. Contact : Feng Liang Life Technologies, a
 division of Invitrogen 9800 Medical Center Drive Rockville
 Maryland 20850, USA Fax : (1) 301 610 8371 Email :
 fliang@lifetech.com URL :
 http://fulllength.invitrogen.com"

BASE COUNT 183 a 232 c 209 g 153 t 1 others
 ORIGIN

alignment_scores:

Quality: 1068.00 Length: 258
 Ratio: 4.395 Gaps: 0
 Percent Similarity: 94.186 Percent Identity: 77.132

alignment_block:

US-09-367-496-8 x AL533854

Align seg 1/1 to: AL533854 from: 1 to: 778

234 GluAlaValTyrArgAlaValThrIleAlaLysGlnAlaAsnCysProLe 250
 2 GAAGCCGTGAATCGTGCATCACCATGCCAACACGACCACTGCCCGCT 51
 250 uTyrValThrLysValMetSerLysGlyAlaAlaAspAlaIleAlaGln 267
 52 GTATATCACCAAGGTGATGAGCAAAARCTCTGTGAGGTGATCGCCAGG 101
 267 lAlaArgArgGlyValValPheGlyGluProIleThrAlaSerLeu 283
 102 CACGGAAGAGGGAACCTGTGTGTATGCGGAGCCCATCACTGCGCCTTC 151
 284 GlyThrAspGlySerHisTyrTrpSerLysAsnTrpAlaLysAlaAla 300
 152 GGAACGACGCGCTCCCATTTACTGAGCAAGAACTGGGCAAGGCTGCTGC 201
 300 aPheValThrSerProProValAsnProAspProThrThrAlaAspHisL 317
 |||||

Possible reversed clone: similarity on wrong strand
Possible reversed clone: poly^r not found
Insert Length: 711 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
unique quality sequence stop: 505

POLYA=NO.

FEATURES
source

```

    .663
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /clone="1048856"
    /clone_lib="Soares_testi
    /sex="male"
    /sex="male"

```

/lab_host="DH10B"
 /note=Vector: pT73D-Pac (Pharmacia) with a modified
 polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
 was prepared from mRNA obtained from Clontech Laboratories
 , Inc., and primed with a Not I - oligo(dT) primer [5',
 TCTTACCAATCTGAAGTGGAGCGGCCCAATTTTTTTTTT 3'].
 Double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Not I and cloned into the Not I
 and Eco RI sites of the modified pT73 vector. Library
 went through one round of normalization to Cot5, and was
 constructed by Bento Soares and M. Fatima Bonaldo.
 204 c 200 g 139 t

	BASE COUNT	ORIGIN
1	0.000	0.000
2	0.000	0.000
3	0.000	0.000
4	0.000	0.000
5	0.000	0.000
6	0.000	0.000
7	0.000	0.000
8	0.000	0.000
9	0.000	0.000
10	0.000	0.000
11	0.000	0.000
12	0.000	0.000
13	0.000	0.000
14	0.000	0.000
15	0.000	0.000
16	0.000	0.000
17	0.000	0.000
18	0.000	0.000
19	0.000	0.000
20	0.000	0.000
21	0.000	0.000
22	0.000	0.000
23	0.000	0.000
24	0.000	0.000
25	0.000	0.000
26	0.000	0.000
27	0.000	0.000
28	0.000	0.000
29	0.000	0.000
30	0.000	0.000
31	0.000	0.000
32	0.000	0.000
33	0.000	0.000
34	0.000	0.000
35	0.000	0.000
36	0.000	0.000
37	0.000	0.000
38	0.000	0.000
39	0.000	0.000
40	0.000	0.000
41	0.000	0.000
42	0.000	0.000
43	0.000	0.000
44	0.000	0.000
45	0.000	0.000
46	0.000	0.000
47	0.000	0.000
48	0.000	0.000
49	0.000	0.000
50	0.000	0.000
51	0.000	0.000
52	0.000	0.000
53	0.000	0.000
54	0.000	0.000
55	0.000	0.000
56	0.000	0.000
57	0.000	0.000
58	0.000	0.000
59	0.000	0.000
60	0.000	0.000
61	0.000	0.000
62	0.000	0.000
63	0.000	0.000
64	0.000	0.000
65	0.000	0.000
66	0.000	0.000
67	0.000	0.000
68	0.000	0.000
69	0.000	0.000
70	0.000	0.000
71	0.000	0.000
72	0.000	0.000
73	0.000	0.000
74	0.000	0.000
75	0.000	0.000
76	0.000	0.000
77	0.000	0.000
78	0.000	0.000
79	0.000	0.000
80	0.000	0.000
81	0.000	0.000
82	0.000	0.000
83	0.000	0.000
84	0.000	0.000
85	0.000	0.000
86	0.000	0.000
87	0.000	0.000
88	0.000	0.000
89	0.000	0.000
90	0.000	0.000
91	0.000	0.000
92	0.000	0.000
93	0.000	0.000
94	0.000	0.000
95	0.000	0.000
96	0.000	0.000
97	0.000	0.000
98	0.000	0.000
99	0.000	0.000
100	0.000	0.000

alignment_scores:		
Quality:	1046.00	Length: 222
Ratio:	4.911	Gaps: 3
Percent similarity:	95.946	Percent Identity: 93.694

alignment_block: AA778518/rev

Align seg 1/1 to reverse of: AA778518 from: 1 to: 663

6 aspAsnPhEAlaLeuIleProGluClyThrAsnGlyIleGluGluArgMe 362
 63 GACAACTTCGGCTGTATCCCGAGGCGACCAACGGCATTCAGAGAGGCAT 614
 62 tSerMetValTrpGluLysCysValAlaSerGlyLYsMetAspGluAsnG 379
 113 GTCGATGTCTGGGAGAATGTGGCTCTGGGAAGATGGACGAGATG 564
 79 lUpheValAlaValThrSerThrAsnAlaAlaLYsIlePheAsnPhEtyr 395
 63 AGTCTGCTCGGTGACCATGCTGCCAAATCTTCAATTTTTTAC 514
 96 ProArgLYsGlyArgValAlaValGlySerAspAlaAspLeuValIleT 412
 113 CCAAAGAAAGCGCGAGTGGCTGTGGGCTCTGCACCTCACTGGTCATAT 464
 112 rPasnProLYsAlaThrLYsIleIleSerAlaLYsThrHisAsnLeuAsn 428
 63 GHAACCCAGGCGCCAAAGATCATCTCTGCCAAGACCCCAATCTGAAC 414
 29 ValGluTrpAsnIlePheGluGlyValGluCysArgGlyAlaProAlaVa 445
 113 GTGGAGTACAAATCTTCGAGGAGTGGAGTCCGGGGAGCCCTGCCGT 364
 445 lValIleSerGlnGlyArgValAlaLeuGluAspGlyLYsMetPheVal 461
 314 GGTCAATAGTCCAGGGCCGAGTGGCGTGGAGCACGGGAAGATGTTTGTCT 314
 478 ThrProGlyAlaGlyArgPheValProArgLYsThrPheProAspPheVa 478
 313 ACCCGGGGGGTGGCGCGCTTCTGCTCCCTCGGAACAACATTCCTCGGACATCTCGT 264
 478 lTrpLYsArgIleLYsAlaArgAsnArgLeuAlaGluIleHisGlyVal 494

seq name: gb est11:AA778518

seq_documentation_block: 24-OCT-2000
 LOCUS 663 bp mRNA EST
 AF0905.1 Soares_testis_NHT Homo sapiens cDNA clone 1048856 3'
 DEFINITION similar to TR:000240 CYTOSOLIC PHOSPHOPROTEIN ; mRNA
 sequence.

ACCESSION AA778518
VERSION AA778518.1
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

SOURCE	Human: Homo sapiens
ORGANISM	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE	1 (bases 1 to 663) GenBank: http://www.ncbi.nlm.nih.gov/ncbiapp .

1 (bases 1 to 663)

AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap/>
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

TITLE
National Cancer Institute
Tumor Gene Index

TOPICAL: Unpublished (1997)
Tumor Gene Index

JOURNAL
COMMENT
Unpublished (1957)
Contact: Robert Strausberg, Ph.D.

CONTACT: ROBERT S. BARNETT, JR.
Email: rsbapbs-re@mail.nih.gov

CDNA Library Preparation: M. Bent

Ph.D.

cdNA Library Arrayed by: Greg Le...

DNA Sequencing by: Washington Un

Clone distribution: NCI-CGAP clone

found through the I.M.A.G.E. Cons

```

263 CTAAGAGGATCAAGCTCGCAACAGGCTGCGGAGAGCTCCACGTCGT 214
|||||
495 ProArgGlyLeuTyrAspGlyProValHisGluValMetValProAlaLys 511
|||||
213 CGCGGTGGGCTGTATGACGGGCGGCTCCACGAGGTGATGCTGCCTGCCAA 164
|||||
511 sProGlySerGlyAlaProAlaAraGlySerCysProGlyLysIleSerV 528
|||||
163 GCCAGGAGCTGCTGCTGCGGCGGCGGCTGCTGCCAGGCAAGATCTCCG 114
|||||
528 alProProValArgAsnLeuHisGlnSerGlyPheSerLeuSerGlySer 544
|||||
113 TCCCTCTGTCGGCAACCTACATCAGTGGGGTTCAGCCTATCTGGTCT 64
|||||
545 GlnAlaAspAspHisIleAlaArgThrAlaGlnLysIleMetAlaPr 561
|||||
63 CAGGCTGATGACCATCGCCCGGAGCAGCAGCAGCAGCATATGCGACC 14
|||||
561 oProGlyGly 564
13 CACTGGCGGC 4

seq_name: gb_est22:AI569716

seq_documentation_block:
LOCUS AI569716 612 bp mRNA EST 29-MAR-1999
DEFINITION tn05a06.x1 NCI-CGAP_Brn25 Homo sapiens cDNA clone IMAGE:2166706 3'
similar to TR:014531 O14531 DIHYDROPYRIMIDINASE RELATED PROTEIN 4.
; mRNA sequence.
ACCESSION AI569716
VERSION AI569716.1 GI:4533090
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 612)
NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute / National Institute of Neurological
Disorders and Stroke, Brain Tumor Genome Anatomy Project
(CGAP/BTGP), Tumor Gene Index
Unpublished (1998)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-re@mail.nih.gov
Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,
Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bbrp/image/image.html
Seq primer: -400P from Gibco
High quality sequence stop: 396.
Location/Qualifiers
1..612
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2166706"
/tissue_lib="NCI-CGAP_Brn25"
/tissue_type="anaplastic oligodendroglioma"
/lab_host="DH10B"
/note="Organ: brain; Vector: pT73D-pac (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTTACCATCTGAATGGAGCGCGCATAGGTTTGTGTTTGTGTTTGTGTTT
T 3']; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT73 vector.
Library is normalized, and was constructed by Bento
Soares and M.Fatima Bonaldo."

```

```

BASE COUNT 115 a 194 c 172 g 130 t 1 others
ORIGIN

alignment_scores:
Quality: 1029.00 Length: 204
Ratio: 5.145 Gaps: 0
Percent Similarity: 98.039 Percent Identity: 97.549

alignment_block:
US-09-367-496-8 x AI569716/rev ..
Align seg 1/1 to reverse of: AI569716 from: 1 to: 612

352 ProGluGlyThrAsnGlyIleGluArgMetSerMetValTrpGluLys 368
|||||
612 CCCGAGTGCACCATCGGCATTGAGGAGCGCATGTCGATGGTCTGGGAGAA 563
|||||
368 sCysValAlaSerGlyLysMetAspGluAsnGluPheValAlaValThr 385
|||||
562 ATGTGTGGCATCTGGGAAGATGGAGAGATGAGTTCGTCGCGGTGACCA 513
|||||
385 erThrAsnAlaAlaLysIlePheAsnPheTyrProArgLysGlyArgVal 401
|||||
512 GTACAAATGCTGCCAAATCTTCATTTTACNCCAAGGAGGCGGAGTG 463
|||||
402 AlaValGlySerAspAlaAspLeuValIleTrpAsnProLysAlaThrLys 418
|||||
462 GCTGTGGCTCTGACGCTGACCTGCTATGGAACCCCAAGGCCACCAA 413
|||||
418 sIleIleSerAlaLysThrHisAsnLeuAsnValGluTyrAsnIlePheG 435
|||||
412 GATCATCTCTGCCAAGACCCACATCTGAACGTGGAGTACAACTCTTCG 363
|||||
435 luGlyValGluCysArgGlyAlaProAlaValIleSerGlnGlyArg 451
|||||
362 AGGAGTGGAGTCCCGGGAGCCCTGCGGTGTCATAGTCAGGCGCGA 313
|||||
452 ValAlaLeuGluAspGlyLysMetPheValThrProGlyAlaGlyArgph 468
|||||
312 GTGGCTGTGGAGAACCGGAGATGTTGTACCCCGGGGCGGCGCTT 263
|||||
468 eValProArgLysThrPheProAspPheValTyrLysArgIleLysAlaA 485
|||||
262 CGTCCCTCGGAAACATCTCCCGACTTTCGTACAGAGGATCAAGCTC 213
|||||
485 rgAsnArgLeuAlaGluIleHisGlyValProArgGlyLeuTyrAspGly 501
|||||
212 GCAACAGGCTGCGGAGATCCACGGTGTGCCCGTGGACTGTATGACGGG 163
|||||
502 ProValHisGluValMetValProAlaLysProGlySerGlyAlaProAl 518
|||||
162 CCGCTCCACGAGGTGATGTCCTGCCAAGCCAGGAGTGGCGTCCGGC 113
|||||
518 aArgAlaSerCysProGlyLysIleSerValProValArgAsnLeuH 535
|||||
112 CCGCGCTCTGCCAAGCAAGATCTCGTCCCTCTGTCGCGCAACCTAC 63
|||||
535 isGlnSerGlyPheSerLeuSerGlySerGlnAlaAspAspHisIleAla 551
|||||
62 ATCAGTCCGGGTTTCAGCTATCTGGGTCTCAGGCTGATGACCATCGCC 13
|||||
552 ArgArgThrAla 555
|||||
12 CGAGCGACAGCA 1

```

seq_name: gb_est51:AW771374

seq_documentation_block:

LOCUS AW771374

DEFINITION hn55d08.x1 NCI-CGAP_Col17 Homo sapiens cDNA clone IMAGE:3027567 3'

similar to SW:DPY4_HUMAN O14531 DIHYDROPYRIMIDINASE RELATED PROTEIN-4 ; mRNA sequence.

EST

mRNA

592 bp

04-MAY-2000


```

alignment_block:
US-09-367-496-8 x AL535160
Align seg 1/1 to: AL535160 from: 1 to: 1049

1 MetSerPheGlnGlyLysSerIleProArgIleThrSerAspArgLe 17
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
192 ATGTCGTACCGAGGCAAGAGAGATCCGCACATCAGGAGTACCCACT 241
17 uLeuLeuArgGlyGlyArgIleValAsnAspGlnSerPheThrAla 34
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
242 CCTCATCAAGGTGGCGGATCATCAAGATGACCAATCCCTTATGCTG 291
34 spValHisValGluAspGlyLeuIleLysGlnIleGlyGluAsnLeu 50
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
292 ACGTCTACCTGGAGATGGAGTCTTATCAACAAATAGGAGAACTTATC 341
51 ValProGlyGlyIleHisThrLeuAspAlaHisCysLeuMetValLeu 67
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
342 GTTCTCTGGTGAAGACCATTTGAAGCCACGCGCGGATGTTATTC 391
67 oGlyGlyValAspValHisThrArgLeuGlnMetProValLeuGlyMet 84
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
392 CGAGGTATTGATGTCACACGATACCTGAGAGCCCTCCAGGGGATGA 441
84 hrProAlaAspPheCysGlnGlyThrLysAlaAlaLeuAlaGly 100
|| |||||:|||||:|||||:|||||:|||||:|||||:|||||:
442 CTGCGCTGATGATCTTCCAGAGGACCGCGGCGGCTGTTGGCGGG 491
101 ThrThrMetIleLeuAspHisValPheProAspThrGlyValSerLeu 117
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
492 ACACGATGATCATTTGACCATGTTCTGTAACCTGGTCCAGCTACT 541
117 uAlaAlaTyrGluGlnTyrArgGluArgAlaAspSerAlaAlaCys 134
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
542 GACCTCTTCGAGAGTGGCAGAGAGCTGACACCAATCTCGTGTG 591
134 spTyrSerLeuHisValAspIleThrArgThrPheHisGluSerIle 150
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
592 ATTACTCTCCAGTGACATCAAGCTGTTACGATGGGTTCCGGAG 641
151 GluLeuGluAlaLeuValLysGlyValAsnSerPheLeuValPh 167
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
642 GAGCTGGAGTGGTGGTGCAGGCAAGGCGTCAATCTTCCAAAGTCTA 691
167 eMetAlaTyrLysArgCysGlnCysSerAspSerGlnMetTyrGlu 184
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
692 CATGCCCTATAAGATGCTTACCAAAATGTCCAGACGACCTCTATGA 741
184 lePheSerIleLeuArgAspLeuGlyAlaLeuAlaGlnValHisAla 200
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
742 CCTTTACCTTCTTAAGGCGCTGGAGCTGTGATCTTGCTCCATGCA 791
201 AsnGlyAspIleValGluGluGlnLysArgLeuLeuGluLeuGly 217
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
792 AATGGAGATTGATAGCTCAGGAACAAAGCGGATCTCGGAGTGGCAT 841
217 eThrGlyProGluGlyHisValLeuSerHisProGluGluValGluAla 234
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
842 CACGGTCCCGAGGCGCATGCTGAGCAGACCTGAAGAGCTGGAGGCG 890
234 LuAlaValTyrArgAlaValThrIleAlaLysGlnAlaAsnCysPro 250
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
891 AGCGGTGTTCGGGCGCATACCATTTGGGGCGGATCACTCCCTCTGT 940
251 TyrValThrLysValMetSerLysGlyAlaAlaAspAlaIleAlaGln 267
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
941 TACATCACCAGGTATGAGCAAGAGTGCAGCGCATCATCGTCTGCG 990
267 alyArgargGlyValValPheGlyGluProfile 279
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
991 CAGGAAGAAAGGGCCCTAGTATTTTGGAGAGCCATTTG 1027

```

seq_name: gb_est29:AL543657

seq_documentation_block:

LOCUS AL543657 1117 bp mRNA EST 16-FEB-2001
 DEFINITION AL543657 LTI_NFL006_PL2 Homo sapiens cDNA clone CS0DI006YB04 5
 prime, mRNA sequence.

ACCESSION AL543657
 VERSION AL543657.1 GI:12876136
 KEYWORDS EST.
 SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 1117)
 Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
 Full-length cDNA libraries and normalization
 Unpublished (2001)

JOURNAL

COMMENT

Contact: Genoscope

Genoscope - Centre National de Sequencage

BP 191 91006 EVRY cedex - France

Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

Location/Qualifiers

1..1117

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="CS0DI006YB04"

/clone_lib="LTI_NFL006_PL2"

/tissue_type="placenta"

/note="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA

was primed with a NotI-oligo(dT) primer. Five prime end

enriched, double-stranded cDNA was digested with Not I and

cloned into the Not I and Eco RV sites of the pCMVSPORT 6

vector. Library was normalized. Library was constructed by

Life Technologies. Contact : Teng Liang Life Technologies,

a division of Invitrogen 9800 Medical Center Drive

Rockville, Maryland 20850, USA Fax : (1) 301 610 8371

Email : fliang@lifetech.com URL :

http://fulllength.invitrogen.com

BASE COUNT 275 a 304 c 288 g 246 t 4 others

ORIGIN

alignment_scores:

Quality: 1015.00 Length: 284

Ratio: 4.028 Caps: 2

Percent Similarity: 88.732 Percent Identity: 67.254

alignment_block:

US-09-367-496-8 x AL543657 ..

Align seg 1/1 to: AL543657 from: 1 to: 1117

289 HistTyrTrpSerLysAsnTrpAlaLysAlaAlaPheValThrSerPr 305

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

1 CATATTGGACAGACAGCTGGGCAAGCGGCTGCTGATTTGTGACATCCC 50

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

305 oProValAsnProAspProThrThrAlaAspHisLeuThrCysLeuLeu 322

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

51 ACCCTGAGCGCTGACCAACTACTCCGGACTACATCACTCTTGTCTGG 100

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

322 erSerGlyAspLeuGluValThrGlySerAlaHisCysThrPheThr 338

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

101 CCAGCGGGGATCTGCAGCTATCTGGGAGTCCCACTGCACCTTCAGCAC 150

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

339 AlaGlnLysAlaValGlyLysAspAsnPheAlaLeuIleProGluGly 355

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

151 GCCCAAGAAAGCAATGGGAAGGACAACTTCACAGCCATTCCTGAGGCG 200

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

355 rAsnGlyIleGluGluArgMetSerMetValTrpGluLysCysValAla 372

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

201 CAATGTGTGGAGGAGCGGATGCTCTCATCTCGGACAAAGCGTGTGGCA 250

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

372 erGlyLysMetAspGluAsnGluPheValAlaValThrSerThrAsnAla 388

DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Plate: LLAM10371 row: 1 column: 11
 High quality sequence stop: 690.
 Location/Qualifiers
 1. 697
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="IMAGE:4502554"
 /clone_lib="NIH_MGC_94"
 /tissue_type="retina"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: eye; Vector: pCMV-SPORT6; Site: 1: NotI; Site 2: SalI; Cloned unidirectionally; oligo-dT primed. Average insert size 3.3 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH_MGC Library."
 169 a 196 c 205 g 127 t

FEATURES
 source

BASE COUNT
 ORIGIN

alignment_scores:
 Quality: 995.00 Length: 235
 Ratio: 4.482 Gaps: 3
 Percent Similarity: 94.468 Percent Identity: 88.511

alignment_block:
 US-09-367-496-8 x BG293352 ..
 Align seg 1/1 to: BG293352 from: 1 to: 697

119 AlAtyrrGluGlnTrpArgGluAlaArgAlaAspSerAlaAlaCysCysAspTy 135
 2 AGCTATGAGCAGTGGCGGAGCGAGCAGACAGCGCGC.TGCTGTGACTA 50
 135 rserLeuHisValaspIlethrargTrpHisGluSerIleLysGluGluL 152
 51 CTCCTTACATGTGACATCTCTCGCTGGCAGCAGAGCACCACCAAGAGAGC 100
 152 euGluAlaLeuValLysGluLysGlyValAsnSerPheLeuValPheMet 168
 101 TGGAGGCCCTAGTCAGGACCAAGGTGTGAACCTCTCTGGTCTTCATG 150
 169 AlaTyrrLysAspArgCysGlnCysSerAspSerGlnMetIrrGluIlePh 185
 151 GCATCAAGGACAGGTGCCAGTGTACTGACGCTCAGATATATGAATCTT 200
 185 eSerIleIleArgAspLeuGlyAlaLeuAlaGlnValHisAlaGluAsnG 202
 201 CA3CCTTCATCCCGGACCTGGGAGCTGTGGCCAGGTGCAGCAGAGAAATG 250
 202 LyaspIleValGluGluGlnLysArgLeuLeuGluLeuGluIleThr 218
 251 GGGACATCTGTGGAGGAGGAGCAAGAGCGCTGCTGGAGCAAGGCATCACT 300
 219 GlyProGluGlyHisValLeuSerHisProGluGluValGluAlaGluAl 235
 301 GGTCTTGAGGGCCATGTGCTGACCCACCAAGAGAGGTAGAGCGCGAGGC 350
 235 aValTyrrArgAlaValThrIleAlaLysGlnAlaAsnCysProLeuTyrr 252
 351 TGTGTACAGACAGTCACTTCCCAAGCAGGCGCAACTGCCCTCTATACG 400
 252 alThrLysValMetSerLysGlyAlaAlaAspAlaIleAlaGlnAlaLys 268
 401 TCACCAAGGTGATGAGCAAGGGTGCAGCTGCATGTGTTGCCCAAGCAAG 450
 269 ArgArgGlyValValValPheGlyGluProIleThrAlaSerLeuGlyTh 285
 451 GCAGGGGGGTGGTGGTCTTCTTGGGGAACCTATCATCTGCCAGCCTGGGCAC 500

251 CAGGAAATATGAGCAAAACAGTGTCTGGCTGTGCAAGCAGCAAAACGCT 300
 389 AlalysIlePheAsnPhetYrProAtqLysGlyArgValAlaValGlySe 405
 301 GCCAAGATCTTCAACCTGTATCCCGCAA.GGAAGAATATCTGTGGTTC 349
 405 rAspAlaAspLeuValIleTrpAsnProLysAlaThrLysIleIleSerA 422
 350 TCACAGCAGCCTCTCATCTGGATCCAGATGTGTGAAGATCCTCTCTG 399
 422 laLysThrHisAsnLeuAsnValGluThrAsnIlePheGluGlyValGlu 438
 400 CCAAGAACCCAGCTCTGGCAGCAGTACACATCTTTGAAGGATGAG 449
 439 CysArgGlyAlaProAlaValIleSerGlnGlyArgValAlaLeuGlu 455
 450 CTGCGCGGCTCTCTGTGTGTCATCTGCCAGGCAAGATCATGCTGGA 499
 455 uAspGlyLysMetPheValThrProGlyAlaGlyArgPheValProArgL 472
 500 AGATGGCAACCTGCACGTGACCCAGGGGGCTGGCCGCTTCATACCTGCA 549
 472 yThrPheProAspPheValTyrrLysArgIleLysAlaArgAsnArgLeu 488
 550 GCCCGTCTCCGACTATGTCTACAGGCGCATTAAGCAGCAGGAGGATG 599
 489 AlaGluIleHisGlyValProArgGlyLeuTyrrAspGlyProValHisG 505
 600 GCACACCTGATGCCCTGCCAGGGGCATGTACGATGGCCCTGTGTGTA 649
 505 uValMetValProAlaLysProGlySerGlyAlaProAlaArgAlaSerC 522
 650 CTTGACCAACCCCAAGGTGGACCCCGCAGGCTCTGTCTGGGGCT 699
 522 ysProGlyLysIleSerValProValArgAsnLeuHisGlnSerGly 538
 700 CTCCTACTCGGCGCAAC...CCACCTGTAAATAATCTTCATGTCGGGA 746
 539 PheSerLeuSerGlySerGlnAlaAspAspHisIleAlaArgArgThrAl 555
 747 TTTAGCCTGTGAGGACCTCAAGTGGATGAGGGGTT...CGCTCAGCCAG 793
 555 aGlnLysIleMetAlaProProGlyGlyArgSerAsnIleThrSerLys 572
 794 CAAGCGCATCTGGCGCCCGCCAGCGCGCTTCTTAATATCATCATCTGA 843
 572 er 572
 844 GT 845

seq_name: gb_est98:BG293352
 seq_documentation_block: 697 bp mRNA EST 21-FEB-2001
 LOCUS BG293352 602390746F1 NIH_MGC_94 Mus musculus cDNA clone IMAGE:4502554 5',
 DEFINITION mRNA sequence.
 ACCESSION BG293352 GI:13052953
 VERSION BG293352.1
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 1 (bases 1 to 697)
 NIH-MGC <http://mgc.nci.nih.gov/>
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-re@mail.nih.gov
 Tissue Procurement: The Cepko Laboratory
 cDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

285 rAspGlySerHisTyrTrpSerLysAsnTrpAlaLysAlaAlaPheV 302
 501 TGATGGCTCACACTACTGAGCAAGAACTGGCCAAAGCTGACGCTTG 550
 302 alThrSerProValAsnProAspProThrAla.AspHisLeuTh 318
 551 TCACATTCACCCCTATCAACCCGGAC...CTACTACTGAGACCACTCAC 597
 318 rCysLeuSerSerGlyAspLeuGlnValThrGlySerAlaHisCysT 335
 598 CTCCTGCTGCTCCAGTGGGACCTCCAGGTCACAGGAGCAGTGC.CACTGCA 646
 335 hrPheThrAlaGlnLysAlaValGlyLysAspAsnPheAlaLeuLe 351
 647 CCTTCAT.ACTGCCCAAGAGGCTGTGGCAAAG...ACAATTCACATGATT 692
 352 Pro 352
 693 CCC 695

seq_name: gb_est30:AUI18714

seq_documentation_block:
 LOCUS AUI18714 773 bp mRNA EST 19-OCT-2000
 DEFINITION AUI18714 HEMBA1 Homo sapiens cDNA clone HEMBA1004213 5', mRNA
 sequence.
 ACCESSION AUI18714
 VERSION AUI18714.1 GI:10933849
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 773)
 AUTHORS Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y., Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and Isogai,T.

TITLE HRI human cDNA project
 JOURNAL Unpublished (2000)
 COMMENT Contact: Takao Isogai
 Genomics Laboratory
 Helix Research Institute
 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
 Tel: 81-438-52-3951
 Fax: 81-438-52-3952
 Email: genomics@hri.co.jp

Research Institute; cDNA library construction: Department of Virology, Institute of Medical Science, University of Tokyo, and Helix Research Institute

FEATURES
 Location/Qualifiers

1..773
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="HEMBA1004213"
 /clone_lib="HEMBA1"
 /tissue_type="whole embryo, mainly head"
 /dev_stage="embryo, 10 weeks"
 /note="vector: pME18SF13"
 BASE COUNT 183 a 213 c 219 g 153 t 5 others
 ORIGIN

alignment_scores:

Quality: 992.50
 Ratio: 4.451
 Percent Similarity: 95.708
 Length: .233
 Gaps: 1
 Percent Identity: 78.112

alignment_block:

US-09-367-496-8 x AUI18714 ..

Align seg 1/1 to: AUI18714 from: 1 to: 773

181 MetTyrGluIlePheSerIleIleArgAspLeuGlyAlaLeuAlaGlnVa 197
 2 ATTTAAGAGTACTGATGGTATCCGGGATATTGGCGCCATAGCCCAAGT 51
 197 IHisAlaGluAsnGlyAspIleValGluGlnLysArgLeuLeuG 214
 52 CCACGAGAAAATGGGACATCATTCACAGAGGAGCAGAGGATCCTGG 101
 214 ILeuGlyIleThrGlyProGluGlyHisValLeuSerHisProGluGlu 230
 102 ATCTGGGCATCACAGCCCGGAGGACATGTGCTGAGCGACCTGAGGAG 151
 231 ValGluAlaGluAlaValTyrArgAlaValThrIleAlaLysGlnAlaAs 247
 152 GTCCGAGGCGGAAGCGTGAATCGTCCATCCATCGCCAAACCCAGACAA 201
 247 cCysProLeuTyrValThrLysValMetSerLysGlyAlaAlaAspAlaI 264
 202 CTGCCCCGTGTATATCACCAAGGTGATGAGCAAAAAGCTCTGCTGAGTCA 251
 264 leAlaGlnAlaLysArgGlyValValValPheGlyGluProIleThr 280
 252 TCGCCAGGCGAGGAAGGAACTGTGGTGTATGGCGAGCCCATCACT 301
 281 AlaSerLeuGlyThrAspGlySerHisTyrTrpSerLysAsnTrpAlaLy 297
 302 GCCAGCTTGGGAGCGGCGCTCCATTTACTGGAGCAAGAACTGGGCCAA 351
 297 sAlaAlaAlaPheValThrSerProValAsnProAspProThrThrA 314
 352 GGCTGCTGCTTGTCTACCTCCCACTTGGAGCCCTGATCCCAACCACTC 401
 314 laAspHisLeuThrCysLeuLeuSerSerGlyAspLeuGlnValThrGly 330
 402 CACACTTCTCAACTCCTGCTGCTGTGGAGACCTNCAGGTCACGGGC 451
 331 SerAlaHisCysThrPheThrAlaGlnLysAlaValGlyLysAspAs 347
 452 AGTGCCCATTTGACGTTTACACTGCCAGAGGCTGTAGGAAGGACAA 501
 347 nPheAlaLeuIleProGluGlyThrAsnGlyIleGluGluArgMetSerM 364
 502 CTTACACCTGATTCGGAGGGCACCATAATGGCACTGAGGAGCGGATGTCG 551
 364 etValTrpGluLysCysValAlaSerGlyLysMetAspGluAsnGluPhe 380
 552 TCATCTGGGACAGGCTGTGGTCACTGGGAGATGGATGAGAACCACTTT 601
 381 ValAlaValThrSerThrAsnAlaAlaLysIlePheAsnPheTyrProAr 397
 602 GTGGCTGTGACCAGCACCAATGCAGNCAAGTCTTCAACCTTTACCCCG 651
 397 gLysGlyArgValAlaValGlySerAspAlaAspLeuValIleTrpAsn 413
 652 GAAAGCCCGCATGCTGTGGGATCCGATGCCGAC.....TGGCAT 691

seq_name: gb_est29:AL537474

seq_documentation_block:

LOCUS AL537474 784 bp mRNA EST 13-FEB-2001
 DEFINITION AL537474 LTI_FLO13_Fbrnl Homo sapiens cDNA clone CS0DF025YN09 5
 prime, mRNA sequence.

ACCESSION AL537474
 VERSION AL537474.1 GI:12800967
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 784)
 AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
 TITLE Full-length cDNA libraries and normalization
 JOURNAL unpublished (2001)

us-09-367-496-8.std.rst

Tue Jul 31 13:08:21 2001

COMMENT

Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: segref@genoscope.cns.fr, web : www.genoscope.cns.fr.

FEATURES

source

1. 784
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="LRI_F1013_FBrnl"
/dev_stage="pooled tissue from post conception fetuses (20 week, 24 week and 26 week)"
/lab_host="DH10B"
/note="organ: Fetal brain; Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and Eco RV sites of the cDNA and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was constructed by Life Technologies. Contact : Feng liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"

BASE COUNT 188 a 236 c 217 g 143 t

ORIGIN

alignment_scores:
Quality: 985.00 Length: 261
Ratio: 4.246 Gaps: 0
Percent Similarity: 88.889 Percent Identity: 68.582

alignment_block:

US-09-367-496-8 x AL537474 ..

Align seg 1/1 to: AL537474 from: 1 to: 784

```

277 GluProIleThrAlaSerLeuGlyThrAspGlySerHisTyrTrpSerly 293
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2 GAGCCATTGCGCCAGCGCTGGGACCGATGGCAGCCATTACTGAGCA 51
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
293 sAsnTrpAlaLysAlaAlaPheValThrSerProProValAsnProA 310
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
52 GAACCTGGGCCAAGGCTGCGCGCTGCTGACCTCCCTCCCTCCCTGAGCCGG 101
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
310 spProThrThrAlaAspHisLeuThrCysLeuLeuSerSerGlyAspLeu 326
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
102 ACCCTACACCGCCGCTACTTACCTCCCTACTGCTGCTGCTGCTGCTGCT 151
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
327 GlnValThrGlySerAlaHisCysThrPheThrThrAlaGlnLysAlaVa 343
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
152 CAGGTCACAGGCGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 201
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
343 lGlyLysAspAsnPheAlaLeuLeuProGluGlyThrAsnGlyIleGluG 360
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
202 GGGCAAGGACCACTTACCTGATCCCGAGGGGTGTCAACGGGATAGAG 251
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
360 luArgMetSerMetValTrpGluLysCysValAlaSerGlyLysMetAsp 376
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
252 ACCGGATGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 301
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
377 GluAsnGluPheValAlaValThrSerThrAsnAlaAlaLysIlePheAs 393
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
302 GAGAACCATGTTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 351
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
393 nPheTyrProArgLysGlyArgValAlaValGlySerAspAlaAspLeuV 410
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
352 CCGTACCAAGGAAAGGCGGATGCGGCTGCGGCTGCGGCTGCGGCTGCGG 401
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
410 allelTrpAsnProLysAlaThrLysIleIleSerAlaLysThrHisAsn 426
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
402 TCATCTGGGACCCCGACAAAGTTGAAGACCATACAGCCCAAAAGTCAAGA 451

```

```

427 LeuAsnValGluTyrAsnIlePheGluGlyValGluCysArgGlyAlaPr 443
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
452 TCGCGCGTGGAGTACAAACATCTTCGAGGGTATGGATGCCAGCGCTCCCC 501
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
443 oAlaValValIleSerGluGlyArgValAlaLeuGluAspGlyLysMetP 460
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
502 ACTAGTGGTTCATCAGCGGCGCAAGATCGTCTTTGAAGACGGAACATCA 551
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
460 heValThrProGlyAlaGlyArgPheValProArgLysThrPheProAsp 476
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
552 ACGTCAACAGGCGATGGGCGCGCTTCATTCGCGGAAGCGTTCCCGGAG 601
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
477 PheValTyrLysArgIleLysAlaArgAsnArgLeuAlaGluIleHisG1 493
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
602 CACCTGTACCGCGCGTCAAAATCAGGAATAGGTTTGGATTGCAAGG 651
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
493 yValProArgGlyLeuTyrAspGlyProValHisGluValMetValProA 510
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
652 GGTTCAGGCGGCGATGATGACGGTCTCTGTAGAGGTACCGAGCTACAC 701
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
510 laLysProGlySerGlyAlaProAlaArgAlaSerCysProGlyLysIle 526
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
702 CCAATATGCAACTCCCGCTCTTCAGCCAAATCTTCGCTTCTTAAACAC 751
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
527 SerValProValArgAsnLeuHisGlnSer 537
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
752 CAGCCCCCACCCTATCAGAAACCTCCACCAGTCC 784

```


.Tue Jul 31 13:08:21 2001

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2001, 11:38:11 ; Search time 23.39 Seconds
(without alignments)
3235.506 Million cell updates/sec

Title: US-09-367-496-8
Perfect score: 2985
Sequence: 1 MSFQGGKIPRITSDRLIR.....RTAQKIMAPGGRSNTLS 572

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL16.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_unclassified.*
- 13: sp_vertebrate.*
- 14: sp_virus.*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2517	84.3	574	13 Q9DDZ6	Q9ddz6 gallus gall
2	1511	50.6	519	11 Q9E0F5	Q9eqf5 mus musculu
3	1475.5	49.4	564	11 Q9JHU0	Q9jhu0 rattus norv
4	1475.5	49.4	564	11 Q9E0F6	Q9eqf6 mus musculu
5	1458.5	48.9	564	4 Q9NQC4	Q9ncq4 homo sapien
6	1453.5	48.7	564	4 Q9NRY9	Q9nry9 homo sapien
7	1450	48.6	563	11 Q9JMG8	Q9jmg8 rattus norv
8	1286.5	43.1	589	5 Q9V3N7	Q9v3n7 drosophila
9	1074.5	36.0	479	2 Q9I676	Q9i676 pseudomonas
10	1043	34.9	531	10 Q9FMP3	Q9fmp3 arabidopsils
11	834.5	28.0	332	5 Q9VND9	Q9vnd9 drosophila
12	692.5	23.2	467	2 Q9E809	Q9e809 streptomyc
13	523.5	17.5	542	3 Q9P903	Q9p903 saccharomyc
14	402.5	13.5	458	2 Q9F465	Q9f465 artthobacte
15	317	10.6	438	2 Q9KAH8	Q9kah8 bacillus ha
16	282	9.4	430	2 Q9EV52	Q9ev52 staphylococ
17	259.5	8.7	422	1 Q9HMH9	Q9hnh9 halobacteri
18	240.5	8.1	428	2 Q9KXR3	Q9kxr3 streptomyc
19	233.5	7.8	428	2 Q9K9V7	Q9k9v7 bacillus ha

20	231.5	7.8	2198	5	Q18990	Q18990 caenorhabdi
21	230.5	7.7	2188	5	Q9VXD5	Q9vxd5 drosophila
22	213	7.1	449	2	Q9PEP0	Q9pep0 xylella fas
23	203	6.8	708	11	Q60504	Q60504 cricetus
24	200	6.7	445	2	Q9HT33	Q9ht33 pseudomonas
25	199	6.7	423	2	Q9F415	Q9f415 pseudomonas
26	199	6.7	499	5	Q9VWM1	Q9vwm1 drosophila
27	192	6.4	474	11	Q35102	Q35102 rattus norv
28	184	6.2	414	1	Q9HIM0	Q9him0 thermoplasma
29	183.5	6.1	409	2	Q50432	Q50432 mycobacteri
30	172	5.8	2275	3	Q93937	Q93937 emericella
31	161	5.4	569	1	Q74030	Q74030 methanobact
32	158.5	5.3	569	1	Q9578	Q9578 methanobact
33	158	5.3	397	2	Q9KQ07	Q9kq07 bacillus ha
34	158	5.3	569	1	Q27599	Q27599 methanobact
35	157.5	5.3	752	5	Q97163	Q97163 drosophila
36	155	5.2	415	2	Q9SLC6	Q9slc6 arthrobacte
37	154	5.2	404	5	Q9NL79	Q9nl79 leishmania
38	153.5	5.1	381	2	Q9KXV7	Q9kxv7 streptomyc
39	153	5.1	431	1	Q9HN51	Q9hn51 halobacteri
40	144.5	4.8	386	2	Q9X247	Q9x247 thermotoga
41	144.5	4.8	402	2	Q9KSO1	Q9ksq1 vibrio chol
42	140.5	4.7	526	1	Q9V2D3	Q9v2d3 pyrococcus
43	136	4.6	569	2	Q9S005	Q9s005 helicobacte
44	134.5	4.5	383	5	Q76138	Q76138 trypanosoma
45	131.5	4.4	445	2	Q9KEV3	Q9kev3 bacillus ha

ALIGNMENTS

RESULT 1
Q9DDZ6 PRELIMINARY; PRT; 574 AA.
AC Q9DDZ6
DT 01-MAR-2001 (TREMREL. 16, Created)
DT 01-MAR-2001 (TREMREL. 16, Last sequence update)
DT 01-MAR-2001 (TREMREL. 16, Last annotation update)
DE COLLAPSEIN RESPONSE MEDIATOR PROTEIN 3.
GN CRMP3.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=RETINA;
RX MEDLINE=20545548; PubMed=10956643;
RA Fukada M., Watakabe I., Yuasa-kawada J., Kawachi H., Kuroiwa A.,
RA Matsuda Y., Noda M.;
RT "Molecular Characterization of CRMP5, a Novel Member of the Collapsin
Response Mediator Protein Family,"
J. Biol. Chem. 275:37957-37965(2000).
DR EMBL; AF249294; AAG37997.1; -;
SQ SEQUENCE 574 AA; 62070 MW; 96E17AC0E96CA5F4 CRC64;

Query Match 84.3% Score 2517; DB 13; Length 574;
Best Local Similarity 81.4%; Pred. No. 1e-173;
Matches 467; Conservative 56; Mismatches 49; Indels 2; Gaps 1;

QY	1	MSFQGGKIPRITSDRLIRGRIVNDQSFYADVHVEDGLIKOIGENLIVPGGIHTIDA	60
Db	1	MSNPGRSGPRCTSEQLLKIGKIVNDQSFYADYIYEDGLIKOIGENLAVPSGRTVDA	60
QY	61	HGLMVLPGGVVHTRLQMPVLGWTADDCCGCTAAAGGTTMILDRVFPDPTGVSLLAA	120
Db	61	YGVLVPGGIDVHTRLQMVGMASADDFYQCTRAALAGGTTMINDHVLPEAGTSLAA	120
QY	121	EQWRERADSAACDYSLHVDIIRWHESKEELEALVKEGVNSFLVFMAYKRCQCSDS	180
Db	121	EQMRQRADSRACCDYALHIDIPRWHESKEELEALVKEGVNSFLVFMAYKDRLQCTDA	180

QY 181 MYEFTSIIRDLGALAOVHAENGDIIVEEQKRLLELIGITGPEGHVLSHPPEVEAEVYRAV 240
 DB 181 MYEFTSIIRDLGALAOVHAENGDIIVEEQKRLLELIGITGPEGHVLSHPPEVEAEVYRAI 240
 QY 241 TIAQANCPPLYVTKVMSKGAADAIAQAKRGVVVFGEPTASLGTDGSHYWSKNWAKAAA 300
 DB 241 TIAQANCPPLYVTKVMSKGAADVAAQAKRGVTVYGEPTASLGADGSHYWSKNWAKAAA 300
 QY 301 FVTSPPVNDPTADHLITCLSSGDLQVTSAGHCTFTTAQKAVGKDNFALIPGNTNGIEE 360
 DB 301 FVTSPPVNDPTADHLITCLSSGDLQVTSAGHCTFTTAQKAVGKDNFALIPGNTNGIEE 360
 QY 361 RWSWYKCVASGKNDENEFVAVTSTNAKIFNFPYPRGRVAVGSDADLVINPKATKII 420
 DB 361 RWSWYKCVASGKNDENEFVAVTSTNAKIFNFPYPRGRVAVGSDADLVINPKATKII 420
 QY 421 SAKTHNLNVEYIFGVEGRCGAPAVVISQGRVALEDGKMFVTPGAGRVPRKTPDFYK 480
 DB 421 SAKTHNLNVEYIFGVEGRCGAPAVVISQGRVALEDGKMFVTPGAGRVPRKTPDFYK 480
 QY 481 RIKARNRLAEHGVPRGLYDGPVHEVMYPAK--FGSAPARASCPGKISVPPVRLHQSG 538
 DB 481 RIKARNRLAEHGVPRGLYDGPVHEVMYPAK--FGSAPARASCPGKISVPPVRLHQSG 538
 QY 539 FSLSGSQADHARTAKOIKINAPPGGRSNITSL 572
 DB 541 FSLSGSQADHARTAKOIKINAPPGGRSNITSL 574
 RESULT 2
 Q9EQF5 PRELIMINARY; PRT; 519 AA.
 ID Q9EQF5
 AC Q9EQF5
 DT 01-MAR-2001 (TREMREL. 16, Created)
 DT 01-MAR-2001 (TREMREL. 16, Last sequence update)
 DE DIHYDROPRIMIDINASE.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SWISS WEBSTER/NIH; TISSUE=WHOLE EMBRYO;
 RX MEDLINE=20545548; PubMed=10956643;
 RA Fukuda M., Watakabe I., Yusa-Kawada J., Kawachi H., Kuroiwa A.,
 RA Matsuda Y., Noda M.;
 RT "Molecular Characterization of CRMP5, a Novel Member of the Collapsin
 RT Response Mediator Protein Family."; J. Biol. Chem. 275:37957-37963(2000).
 DR EMBL: AF249296; AAG37999.1; -.
 SQ SEQUENCE 519 AA; 56681 MW; F2C53836B2120ABD CRC64;
 Query Match 50.6%; Score 1511; DB 11; Length 519;
 Best Local Similarity 58.3%; Pred. No. 4.7e-101;
 Matches 287; Conservative 66; Mismatches 135; Indels 4; Gaps 1;
 QY 16 RLIRGGVINDDQSFYADVHVEDGLIKQIGENLIVP-----GGIHTTDAHGLMVLPGVD 71
 DB 6 RLIRGGVINDDQSFYADVHVEDGLIKQIGENLIVP-----GGIHTTDAHGLMVLPGVD 71
 QY 72 VTRLOMPVLGMPADDFCQGTCAALAGGTTMILDHVPDGTGVSLLAAYEQRERADSA 131
 DB 66 THTHMQFFMGSQSDVDFYQGTCAALAGGTTMILDHVPDGTGVSLLAAYEQRERADSA 131
 QY 132 CCDSYSLHVDITRWHESTKEELALVKEGVNSFLVFMAYDRQCSDSQMVEYFSTIRDL 191
 DB 126 CCDSYSLHVDITRWHESTKEELALVKEGVNSFLVFMAYDRQCSDSQMVEYFSTIRDL 191
 QY 192 GALAQVHAENGDIIVEEQKRLLELIGITGPEGHVLSHPPEVEAEVYRAVIAQANCP 251
 DB 192 GALAQVHAENGDIIVEEQKRLLELIGITGPEGHVLSHPPEVEAEVYRAVIAQANCP 251

DB 186 GAIAOVHAENGDLIAEGAKKMLALGITGPEGHELCRPAVEAEATLRAITIASAVNCP 245
 QY 252 VTKVMSKGAADAIAQAKRGVVVFGEPTASLGTDGSHYWSKNWAKAAAFVTSPPVNDP 311
 DB 246 VVHVMKSAKVVADARRAGNVVYGEPTAAGLGTGQYWESESHAAHHVWGPPLRDP 305
 QY 312 TTADHLITCLSSGDLQVTSAGHCTFTTAQKAVGKDNFALIPGNTNGIEEBSMWKCV 371
 DB 306 LTPGFLMDLLANGDLTTGSDNCTFNCQKALGKDDFTKIPNGVNGVEDRMSVWEKGV 365
 QY 372 SGKMDENEFVAVTSTNAKIFNFPYPRGRVAVGSDADLVINPKATKIIISAKTHNLNVEY 431
 DB 366 SGKMDENEFVAVTSTNAKIFNFPYPRGRVAVGSDADLVINPKATKIIISAKTHNLNVEY 431
 QY 432 NIFEGVRCGAPAVVISQGRVALEDGKMFVTPGAGRVPRKTPDFYKRIKARNRLAEI 491
 DB 426 NIFEGVRCGAPAVVISQGRVALEDGKMFVTPGAGRVPRKTPDFYKRIKARNRLAEI 491
 QY 492 HGVPRLYDGPV 503
 DB 486 VPKRAPHYKGEV 497
 RESULT 3
 Q9JHU0 PRELIMINARY; PRT; 564 AA.
 ID Q9JHU0
 AC Q9JHU0
 DT 01-OCT-2000 (TREMREL. 15, Created)
 DT 01-OCT-2000 (TREMREL. 15, Last sequence update)
 DE ULIP-LIKE PROTEIN
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Horiuchi M., Betz H.;
 RT "Ulup and dihydropyrimidinase like protein."; Submitted (DSC-1998) to the EMBL/GenBank/DBJ databases.
 RL EMBL: AJ31436; CAB95193.1; -.
 DR InterPro; IPR002195; -.
 DR Pfam; PF00744; Dihydroorotase; 1.
 SQ SEQUENCE 564 AA; 61540 MW; E078B9002F54975E CRC64;
 Query Match 49.4%; Score 1475.5; DB 11; Length 564;
 Best Local Similarity 50.2%; Pred. No. 2e-98;
 Matches 281; Conservative 95; Mismatches 173; Indels 11; Gaps 3;
 QY 14 SDRLLIRGGVINDDQSFYADVHVEDGLIKQIGENLIVPGLIHTDAHGLMVLPGVDVH 73
 DB 7 SVRLIKGGKVVNDCTHEADVVIENGIIQQVRELMPGGAKVIDATGKLIVPGIDTS 66
 QY 74 TRLOMPVLGMPADDFCQGTCAALAGGTTMILDHVPDGTGVSLLAAYEQRERADSA 133
 DB 67 THTHMQFFMGSQSDVDFYQGTCAALAGGTTMILDHVPDGTGVSLLAAYEQRERADSA 133
 QY 134 DYSLHVDITRWHESTKEELALVKEGVNSFLVFMAYDRQCSDSQMVEYFSTIRDLGA 193
 DB 127 DYSLHVDITRWHESTKEELALVKEGVNSFLVFMAYDRQCSDSQMVEYFSTIRDLGA 193
 QY 194 LAQVHAENGDIIVEEQKRLLELIGITGPEGHVLSHPPEVEAEVYRAVIAQANCP 253
 DB 187 IPVHAENGDIIVEEQKRLLELIGITGPEGHVLSHPPEVEAEVYRAVIAQANCP 253
 QY 254 KVMKGAADAIAQAKRGVVVFGEPTASLGTDGSHYWSKNWAKAAAFVTSPPVNDP 313
 DB 247 NVSSISAGDVIAAKMQCKVLAETNAHATLTGUYTHQMSHAAAYVTPPLRLDNT 306
 QY 314 ADHLITCLSSGDLQVTSAGHCTFTTAQKAVGKDNFALIPGNTNGIEEBSMWKCV 373
 DB 307 STYLSLLANDTLNIVASHRPFTHKQMGKEDFTKIPNGVNGVEDRMSVWEKGV 366

```

374 KMDNEFVAVTSTNAKIFNFPYPRKGRVAVGSDADLVINWPKATKIISAKTHNLNVEYNI 433
367 KMDNEFVAVTSSNAKILNLYPRKGRITPGADADVVDVWDPEATKTIISASTQVGGDFNL 426
434 FEGVECRGAPAVVISOGVALEDGKMFVTPGAGRFVPRKTFPDPFVYKRIKARNRLAEIHG 493
427 YENMRCHGVPLVITSRGVYVYENGVMCAEGTKFCPLRSFPDIVYKLVQREKTLKVRG 486
494 VPRGLYDGPVHEVMVPAKPGSGAPARASCPCPKISVPPV-----RNHQSFGSLSGSQAD 547
487 VDRTPYLDGVAIVNPGKEMGTPL-ADTPTR----PVTRHGGMRDLHESFSLSGSQID 541
548 DHIAARRTAQKIMAPPGGRSN 567
542 DHVPRASARILAPPGGRSS 561

RESULT 4
Q9EQF6 PRELIMINARY; PRT; 564 AA.
AC Q9EQF6 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DE 01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE COLLAPSED RESPONSE MEDIATOR PROTEIN 5.
GN CRMP5.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SWISS WEBSTER/NIB; TISSUE=WHOLE EMBRYO;
RX MEDLINE=20345548; PubMed=10956643;
RA Fukuda M., Watakabe I., Yuasa-Kawada J., Kawachi H., Kuroiwa A.,
RA Matsuda Y., Noda M.;
RT "Molecular Characterization of CRMP5, a Novel Member of the Collapsin
RT Response Mediator Protein Family.";
RL J. Biol. Chem. 275:37957-37965(2000).
DR EMBL; AF249295; AAG37998.1; -.
SQ SEQUENCE 564 AA; 61516 MW; CA93790FC8F9CD98 CRC64;

Query Match 49.4%; Score 1475.5; DB 11; Length 564;
Best Local Similarity 50.2%; Pred. No. 2e-98;
Matches 281; Conservative 95; Mismatches 173; Indels 11; Gaps 3;

QY 14 SDRLIRGRIVNDQSFYADVHVEDGLIKQIGENLIVPGGIHTIDAGLMLVPGGVVH 73
DB 7 SVRLIKGGKVVNDCTHEADVIESGIIQVGRLEMPGAKVIDATGKLVIPGGIDTS 66
QY 74 TRLOMPVLGTPADDFCOGTKAALAGGTTMLDHPVDTGVSLLAAVEQWRERADSACC 133
DB 67 THFHQTFMNAATCVDVDFYHGTKAALVGGTTMIIGHVLPDKETSLVAYEKCRALADPKVCC 126
QY 134 DYSLHVDITRWHEISKEELEALYKEGVNSFLVPMAYKDRCCSDSOMYEIESTIRDLGA 193
DB 127 DYALHVGITWPAKVAEMETLVREKGVNSFQMFMTYKDLMLRDSLEYQVHFACRDIGA 186
QY 194 LAQVHAENGDIVEEQKRLLELGTGPEGHVLSHPPEEVAEAVYAVYTIKQANCPLYVT 253
DB 187 IPRVHAENGELVAEGAKALDLGITGPEGIEISHPEELEAEATHRVITIANRTHCPILYV 246
QY 254 KVMKGAADAIAQAKRRGVVFGEPITASLTGDSHWSKNWAKAAAFVSPVNPDPPT 313
DB 247 NVSSISAGDVIAAKQKGVVLAETTTAHTATLTGLHYHODMSHAAAYVTPPLRLDNT 306
QY 314 ADHLTCLSSGDLQVTSAGHCTFTTAQKAVGKDNFALIPETNGTEERMSVMWKECVASG 373
DB 307 STYLSLANDTLNIVASDRHPFTTKQAMGKEDFTKPHGVSGVQDRMSVWVERGVGG 366
QY 374 KMDNEFVAVTSTNAKIFNFPYPRKGRVAVGSDADLVINWPKATKIISAKTHNLNVEYNI 433

```

```

367 KMDNEFVAVTSSNAKILNLYPRKGRITPGADADVVDVWDPEATKTIISASTQVGGDFNL 426
434 FEGVECRGAPAVVISOGVALEDGKMFVTPGAGRFVPRKTFPDPFVYKRIKARNRLAEIHG 493
427 YENMRCHGVPLVITSRGVYVYENGVMCAEGTKFCPLRSFPDIVYKLVQREKTLKVRG 486
494 VPRGLYDGPVHEVMVPAKPGSGAPARASCPCPKISVPPV-----RNHQSFGSLSGSQAD 547
487 VDRTPYLDGVAIVNPGKEMGTPL-ADTPTR----PVTRHGGMRDLHESFSLSGSQID 541
548 DHIAARRTAQKIMAPPGGRSN 567
542 DHVPRASARILAPPGGRSS 561

RESULT 5
Q9NOCA PRELIMINARY; PRT; 564 AA.
AC Q9NOCA 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE HYPOTHETICAL 61.4 KDA PROTEIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=SPINAL CORD;
RA Horiuchi M., Betz H.;
RT "Human homologue of Ulip6.";
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ251275; CAB95124.1; -.
DR InterPro; IPR002195; -.
DR Pfam; PF00744; Dihydroorotase; 1.
KW Hypothetical protein.
SQ SEQUENCE 564 AA; 61394 MW; F370DD35B0B49935 CRC64;

Query Match 48.9%; Score 1458.5; DB 4; Length 564;
Best Local Similarity 49.5%; Pred. No. 3.3e-97;
Matches 277; Conservative 100; Mismatches 172; Indels 11; Gaps 3;

QY 14 SDRLIRGRIVNDQSFYADVHVEDGLIKQIGENLIVPGGIHTIDAGLMLVPGGVVH 73
DB 7 SVRLIKGGKVVNDCTHEADVIESGIIQVGRLEMPGAKVIDATGKLVIPGGIDTS 66
QY 74 TRLOMPVLGTPADDFCOGTKAALAGGTTMLDHPVDTGVSLLAAVEQWRERADSACC 133
DB 67 THFHQTFMNAATCVDVDFYHGTKAALVGGTTMIIGHVLPDKETSLVAYEKCRALADPKVCC 126
QY 134 DYSLHVDITRWHEISKEELEALYKEGVNSFLVPMAYKDRCCSDSOMYEIESTIRDLGA 193
DB 127 DYALHVGITWPAKVAEMETLVREKGVNSFQMFMTYKDLMLRDSLEYQVHFACRDIGA 186
QY 194 LAQVHAENGDIVEEQKRLLELGTGPEGHVLSHPPEEVAEAVYAVYTIKQANCPLYVT 253
DB 187 IARVHAENGELVAEGAKALDLGITGPEGIEISRPEELEAEATHRVITIANRTHCPILYV 246
QY 254 KVMKGAADAIAQAKRRGVVFGEPITASLTGDSHWSKNWAKAAAFVSPVNPDPPT 313
DB 247 NVSSISAGDVIAAKQKGVVLAETTTAHTATLTGLHYHODMSHAAAYVTPPLRLDNT 306
QY 314 ADHLTCLSSGDLQVTSAGHCTFTTAQKAVGKDNFALIPETNGTEERMSVMWKECVASG 373
DB 307 STYLSLANDTLNIVASDRHPFTTKQAMGKEDFTKPHGVSGVQDRMSVWVERGVGG 366
QY 374 KMDNEFVAVTSTNAKIFNFPYPRKGRVAVGSDADLVINWPKATKIISAKTHNLNVEYNI 433
367 KMDNEFVAVTSSNAKILNLYPRKGRITPGADADVVDVWDPEATKTIISASTQVGGDFNL 426

```

Qy	434	PEGVCEKGAFAVVISQGRVALEDGKMFVTPGAGRFVPRKTFDPFVYVKRTKARNLAEITHG	493
Db	427	YENMRCHGVPLVITISRGVYVYENGVEWCAEGTGKFCPLSFDPDYKKLVQREKTLKVRG	486
Qy	494	VPRGLYDGPVHEVWVPAKPCSGAPARAPACPGKISVPPV-----RNLHQSGFSLSGSQAD	547
Db	487	VDRTPYLGDVAVVHPGCKEMGFPPL-ADTPTR----PVTRHGGMRDLHSSFSLSGSQID	541
Qy	548	DHIARRTAQKIMAPPGGRSN	567
Db	542	DHVPKRA SARILAPPGGRSS	561
RESULT	6		
Q9NRY9			
ID	Q9NRY9	PRELIMINARY;	PRT; 564 AA.
AC	Q9NRY9		
DT	01-OCT-2000	(TrEMBLrel. 15, Created)	
DT	01-OCT-2000	(TrEMBLrel. 15, Last sequence update)	
DE	01-MAR-2001	(TrEMBLrel. 16, Last annotation update)	
DE	COLLAPSN RESPONSE MEDIATOR PROTEIN-5.		
GN	CRMP5.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
NCBI_TaxID	9606;		
RP	[1]		
RP	SEQUENCE FROM N.A.		
RA	Yu Z, Kryzer T.J., Griesmann G.E., Lennon V.A.;		
RT	Collapsin response mediator protein-5 (CRMP5), a novel autoantigen		
RT	associated with paraneoplastic neurological disorders.;		
RL	Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.		
RL	EMBL; AF157634; AAF80348.1; -		
DR	InterPro; IPR002195; -		
DR	Pfam; PF00744; Dihydroorotase; 1.		
SEQUENCE	564 AA; 61459 MW; 949947A983735340 CRC64;		

Query Match 48.7%; Score 1453.5; DB 4; Length 564;
Best Local Similarity 49.3%; Pred. NO. 7.6e-97;
Matches 276; Conservative 100; Mismatches 173; Indels 11; Gaps 3;

Qy	14	SDRLLIRGTVNDQSFYADVHVHEDGLIKQIGENLIVPGGIHTIDAHGLMVLPGGVDDH	73
Db	7	SVRIIRKGGKVNDCTHEADVYIENGIIQOVGRELMIPGSAKVIDATGKLVIPGGIDHS	66
Qy	74	TRLQMPVLVGMTPADDFCCQTKAALAGTTMILDHVFPDTCVSLAAYEQWRERADSAACC	133
Db	67	THFHQTFMNATCVDDFYHGTRKAALVGTTMILGHVLPDKETSLVDAYEKCRGLADPKVC	126
Qy	134	DYSLHVDITRWHSIEKEELEALYKRGVNSFLVMAYKORCCSDSOMYEIIFSIIIRDLCA	193
Db	127	DYALHVGITWAPKVKRAEMETLVREKGVNSQFMETYKDLMYLRDSELQVVLHACKDKCA	186
Qy	194	LAOVHAENGDIVEEOKRLELIGITGPEGHVLSHPVEEAAYVAVTIAQANCPLYVT	253
Db	187	IARVHAENGELVAGAKAELDLGITGPEITISRPELEAEATHRVITIANTHGPIYLV	246
Qy	254	KVMSKGAADAIQAKRRGVVVFGPEITASLIGTDGSHYKSNKAKAAAFVTSPPVNDPPT	313
Db	247	NVSSIAGDVIAAKMQGVLAETTTAHATLTLGHYYHQDMSHAAAYVVPPLRDNVT	306
Qy	314	ADHLTCLLSGGDLQVNGSAHCTFTTAQKAVGKDNFALPEGTVNGIEERMSYMWKCVASG	373
Db	307	STYLSMLLNDLTNIIVASDHRPFTTKQAMGKDEFTKIPHGVSQGVQDRMSYIWERGVVGG	366
Qy	374	KMDNEFEVAVTSTNAKIPENFYPRKGRVAVGSDADLVINWPKATKIISAKTHNLNVEYNI	433
Db	367	KMDNEFEVAVTSSNASKLNLNYPKGRIIIPGAHADVVMDPEATKTIISASTQVQGGDFNL	426
Qy	434	FEQVCECRGAPAVVISQGRVALEDGKWFYTPGAGRFVPKRTTPDFVKKRIKARNLAEIHG	493
Db	427	YENMRCHGVPLTISGRVYVYENGVMCAEGTGKFLRSPDVTYKVLQREKTLKVRG	486

QY	494	VPRGLYDGPVHEVMVPKPGSGAPARASCPGKISVPV-----RNHQSFGSLSGSQAD	547
D6	487	VDRTPYLGDVVAVVHPGGKKEMGTPL-ADIPTR----PVTRHGGMRLDHSSFSLSGSQID	541
QY	548	DHIARTTAQKIMAPPGRSN	567
D6	542	DHVFKRASRILAPPPGRSS	561
RESULT	7		
Q9JMG8			
ID	Q9JMG8	PRELIMINARY;	PRT; 563 AA.
AC	Q9JMG8;		
DT	01-OCT-2000 (TrEMBLrel. 15, Created)		
DT	01-OCT-2000 (TrEMBLrel. 15, Last sequence update)		
DE	01-MAR-2001 (TrEMBLrel. 16, Last annotation update)		
DE	DIHYDROPYRIMIDINASE-RELATED PROTEIN.		
OC	Rattus norvegicus (Rat).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.		
OX	[NCBI_Taxid=10116;		
RNA	[1]		
RP	SEQUENCE FROM N.A.		
RA	Yanagi S., Inatome R., Yamamura H.;		
RT	"Dihydropyrimidinase-related protein.";		
RRL	Submitted (JUN-1999) to the EMBL/GenBank/DDBJ databases.		
EMBL;	AB029432; BAAB9475.1; "		
InterPro:	IPR002195; -		
DR	Pfam: PF00744; Dihydroorotase; 1.		
SEQUENCE	563 AA; 61395 MW; B36767BF14A0004D CRC64;		

Query Match	48.68;	Score 1450;	DB 11;	Length 563;
Best Local Similarity	49.88;	Pred. No. 1.4e-96;		
Matches 279;	Conservative	94;	Mismatches 175;	Indels 12;
				Gaps 4;

Qy	14	SDRLLRGRIVNDQSFYADVHVHVEDGLIKQIGENILVPGGIHTIDAHGLMWLPGGVDVH	73
Db	7	SVRLIKGGVNVNDCDTHADAYIENGIIQOVGRELWIPGGAKVIDATGKLVIPGGIDTS	66
Qy	74	TRLQMPVLGMITPADDTCQGTKAALAGGTTMILDHVPDTCVSLLAAYEQWRERADSAAOC	133
Db	67	THFQTFMNATCYDDFVHGTKAALVGTTMILGHVLPDKETSLVEAYEKRALADPKV-C	125
Qy	134	DYSLHVDITRWHSIEKBEALVKYKGVNSFLVFMAYKDRQCSDSQMYEIFSIIIRDLGA	193
Db	126	DYALHVGITWAPKVKRAEMETLVREKGVNSQFMFTYKOLYMLRDELQVPHACRDFGA	185
Qy	194	LAQVHAENGDIVEEQKRLELIGTICPEGHVLSHPHEVEAEAVYRAVTTAKQANCPLYVT	253
Db	186	IPRVHAENGELVAEGAKEALDLGTCPEGIEISHPHELEAEATHRVITIANTHCPIYLV	245
Qy	254	KVMSKGAADAAIAQAKRGVVVVEEPTTASLGTDGSHYKSNKAKAAAFVTSPPVNDPPT	313
Db	246	NVSSISAGDIVAAKMOGVKVLAEETNAHATITGLHYHQDWSHAAAYVTVPEPLRLDNT	305
Qy	314	ADHLTCLSSGDLQVTSACHTFTTAAQKAVGKDNFALPEGTNGTIEERSMWYEWKCVASG	373
Db	306	STYLSLLANDTLNIVASDRPFTTKQKANGHEDTKTPHGVSQVDRMSVWYEWGVVGG	365
Qy	374	KMDENEFVAVTSNAAKIFNFYPRGRVAVGSDADLWVNPRAKTIISAKTHNLNVEYNI	433
Db	366	KMDENRFVAVTSNAAKILNLYPRGRIIPGADADAVVWMDPEATKTIISASTQVQGGDFNL	425
Qy	434	FEQVECRGAPAVVISQGRVVALEDGKMFVTPGAGREVPRKTPFDVFKYKIKARNLAEIHG	493
Db	426	YENMRCHGVPLVTISGRKVYVYENGVEFMCAGETGNFCPLRSRPFYVKKLVOREKTLKRG	485
Qy	494	VPRGLYDGPVHEVMYPAKPGSCAPARASCPKISVPPV-----RNLHQSGFSLSGSQAD	547
Db	486	VDRTPYLGDAVVVNPGRKMGMTPL-ADTPFR----PVRHGGRMDLHESFSLSGSQID	540

QY	548	DHIARRTAQKIMAPPGRSN	567	DR	EMBL; AF160900; AAD46840.1; -
Db	541	DHVPKRASARILAPPGGRSS	560	DR	EMBL; AF077837; AAD52007.1; -
				DR	FlyBase; FBgn0023023; CRMP.
				DR	InterPro; IPR002195; -
				DR	Pfam; PF00744; Dihydroorotase; 1.
				SQ	SEQUENCE: 589 AA; 64430 MW; 959A44B657CF3891 CRC64;
RESULT	8				
Q9V3N7		PRELIMINARY;	PRT; 589 AA.		
ID	Q9V3N7				
AC	Q9V3N7				
DT	01-MAY-2000	(TrEMBLrel. 13, Created)			
DT	01-MAY-2000	(TrEMBLrel. 13, Last sequence update)			
DT	01-MAR-2001	(TrEMBLrel. 16, Last annotation update)			
DE	CRMP PROTEIN				
GN	CRMP OR BCDA:HL02693 OR CG1411.				
OS	Drosophila melanogaster (Fruit fly).				
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;				
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;				
OC	Ephydroidea; Drosophilidae; Drosophila.				
OX	NCBI_TaxID=7227;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN-BERKELEY;				
RA	MEDLINE-20196006; PubMed-10731132;				
RA	Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,				
RA	Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,				
RA	George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,				
RA	Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,				
RA	Brandon R.C., Rogers J., Blazer R.G., Champe M., Pfeiffer B.D.,				
RA	Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,				
RA	Abriel J.F., Agbayani A., An H.-J., Andrews-Pfannkoch L., Beasley E.M.,				
RA	Baillet R.M., Basu A., Baxendale J., Bayraktaroglu L., Bhandari D.,				
RA	Beeon K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,				
RA	Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,				
RA	Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,				
RA	Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,				
RA	de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,				
RA	Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,				
RA	Dubin K.J., Evangelista C.C., Ferraz C., Kravitz S., Kulp D., Lai Z.,				
RA	Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,				
RA	Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,				
RA	Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,				
RA	Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,				
RA	Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,				
RA	Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,				
RA	Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,				
RA	Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,				
RA	Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,				
RA	Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,				
RA	Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,				
RA	Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,				
RA	Palazzo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,				
RA	Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,				
RA	Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,				
RA	Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,				
RA	Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,				
RA	Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,				
RA	Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,				
RA	Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,				
RA	Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,				
RA	Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;				
RT	"the genome sequence of Drosophila melanogaster.";				
RL	Science 287:2185-2195(2000).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RA	Tsang G., Brokstein P., Frise E., Harvey D., Evans-Holm M.,				
RA	Lewis S.E., Suh C., Rubin G.M.;				
RT	"Full Length Drosophila melanogaster cDNA sequence.";				
RL	Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.				
RN	[3]				
RP	SEQUENCE FROM N.A.				
RA	Wang L.H., Strittmatter S.M.;				
RT	"Drosophila CRMP is expressed in the developing nervous system.";				
RL	Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.				
EMBL	AE003602; AAF52002.1; -				

DR	EMBL; AF160900; AAD46840.1; -	Query Match	43.1%;	Score	1286.5;	DB	5;	Length	589;
DR	EMBL; AF077837; AAD52007.1; -	Best Local Similarity	43.9%;	Pred. No.	9.7e-85;				
DR	FlyBase; FBgn0023023; CRMP.	Matches	266;	Conservative	92;	Mismatches	175;	Indels	73;
DR	InterPro; IPR002195; -								
DR	Pfam; PF00744; Dihydroorotase; 1.								
SQ	SEQUENCE: 589 AA; 64430 MW; 959A44B657CF3891 CRC64;								
QY	7	KSIP---RITSDBLLIRGGRIIVNDOSFYADVHVDEGLIKOIG--ENLIVPGGIHTIDAH	61						
Db	9	KKYPPIHQSAQNRYIKNGEIVNHDKSFADYVIEDGIKIFVGPSSIEITPGGVRTIDS	68						
QY	62	GLMVLPGGVVHTRLQMPVLGMPADDFCOGTAKAALAGGTTMILDHVPDPTGVSLAAAYE	121						
Db	69	GLMIIPGGIDPHTHMQLPFGGAVAVDDFYHGTRKAAVAGGTTMIIDFVLPKHESMIEAYD	128						
QY	122	QWRERADSAACCDYSLHVDITRWGHESIKKELEALVKEGVNSFLVFMAYKDRQCQSDSM	181						
Db	129	KWRSWADPKVKCCDYGLHVGITWMSKVSSEEIGILCKELGVNSFKTFMAYKGLQLNDSL	188						
QY	182	YEIFSIIIDLALCAQVHAENGDIVEEQKRLLELIGITGPEGHVLSHPPEVEAEAVYRAVT	241						
Db	189	LDVFERIHLNGVAMVHAENGDIITAKNTQRLLAEGINGEGHELSPPEVEAEAVHRACV	248						
QY	242	IAQANCPLYVTKVMSKGAADAIQAOKRGVVVFGEPTASLGTD-----GSHVWSKNW	295						
Db	249	LAHQMKTPLFVSGLTSSAELVGRARRSGYCVFGETLASSLGRSMSAVPKGDRIIYA---	305						
QY	296	AKAAAFVTSPPVNDPTADHLTCLLSGDDLVQVTSAGHCTTTAKAVKGVKDNFALIEGT	355						
Db	306	-----ITSPPIRESAETPRQLMKSALVDDQLGLGSDNCTFNKEHKGALGDKTIPNGV	359						
QY	356	NGIERMSVWEKCVASGKMDENEFVAVTSTNAKIFNFYPRKRVAVGSDADLVINPK	415						
Db	360	NGVEDMSLVWEKGVHAGLLDPCRFVAVTSTNAKIFNFYPRKRVAVGSDADLVINPN	419						
QY	416	APKIIISAKTHNLVNYNIFEGVECGAPAVVISQGRVALEDGKMFVTPGAGRVPRKTFP	475						
Db	420	APRTISKTHHACDFNIFEGMTVHGVCDFVLVRGICAEGRNVAVAGFGFFITPVPR	479						
QY	476	DIYVKRIKAR-----NRLAEIHGVPRGLYDGVVHEVWVPAKPGSGAP	517						
Db	480	PHYDILIEGKVQSQPEEQHEEKQNGSMARFAEL-----DIQIPVQ-----EP	522						
QY	518	ARASCPGKITSVPP-----VRLHQSGESLSGSOADDHIAFRTAOKINAP	561						
Db	523	IGAMLAGNLAMPAGSLCSTPSVGRVGRVQDKRLQESSFSI---SEELDRSGVRAKIVKNP	580						
QY	562	PGGRSN	567						
Db	581	PGKSS	586						
RESULT	9								
Q91676		PRELIMINARY;	PRT; 479 AA.						
ID	Q91676								
AC	Q91676								
DT	01-MAR-2001	(TrEMBLrel. 16, Created)							
DT	01-MAR-2001	(TrEMBLrel. 16, Last sequence update)							
DT	01-MAR-2001	(TrEMBLrel. 16, Last annotation update)							
DE	DIHYDROPYRIMIDINASE.								
GN	PA0441.								
OS	Pseudomonas aeruginosa.								
OC	Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;								
OC	Pseudomonas.								
OX	NCBI_TaxID=287;								
RN	[1]								
RP	SEQUENCE FROM N.A.								
RC	STRAIN-PA01;								

RX MEDLINE-204373737; PubMed-10984043;
 RA Stover C.K., Pham X.-O.T., Erwin A.L., Mizoguchi S.D., Warren P.,
 RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
 RA Garber R.L., Golltry L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
 RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
 RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
 RA Reiser J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.,
 RT "Complete genome sequence of *Pseudomonas aeruginosa* PA01, an
 opportunistic pathogen."
 RL Nature 406:959-964(2000).
 DR EMBL; AE004481; AAG03830.1; --
 DR InterPro; IPR002195; --
 DR Pfam; PF00744; Dihydroorotase; 1
 SQ SEQUENCE 479 AA; 52212 MW; 46582DFE22E38925 CRC64;

Query Match 36.0%; Score 1074.5; DB 2; Length 479;
 Best Local Similarity 45.8%; Pred. No. 1.5e-69;
 Matches 216; Conservative 81; Mismatches 166; Indels 9; Gaps 4;
 QY 17 LLIRGRVNDQSFYADVHVEDGLIKOIGENLIVPGGIHTIDAGLMLVPGGVHTRL 76
 Db 3 LLIRGATVTVTHEESYRADVLCANGLIQAIGENLETSPGCDVLDGGGYLMPGGIDPHTM 62
 QY 77 QMPVLGMPADDFCOGTTKAALAGCTTMLDHFVDPDTGVSLLAAYEQWRERADSACDYS 136
 Db 63 QLPFGVIVASEDFSGTAAGLAGCTTSLIDFVIPNPRQSLLEAFHTWRGWAQKSA-ADYG 121
 QY 137 LHVDITRWHESIKEELALYKEGVNSFLVFMAYKDCQSDSQSOMYEIIFIRDLGALAQ 196
 Db 122 FHVAITWMSDEVAREMGELVAGHGVNSFKHFMAYKKNMAADOTLVASFCELEAGVPT 181
 QY 197 VHAENGDIIVEEOKRLELGTGPEGHVLSHPVEAEAVYRAVTTIAKQANCPYITKVM 256
 Db 182 VHAENGELVFLHQLKLAQGLTGPFAHLSRPPQVEGAASRAIRAEITGLTPLYLHIS 241
 QY 257 SKGAADIAQAKRGVVVFGEPITASLTGDSHYWKNWAKAAAFVTSPPVNDPTADH 316
 Db 242 SREALDEIAYARAKGPVYCEVLGHLHLLDSDSVYRHPDWAATAGYVNSPPFRP- 297
 QY 317 LTCL--LSSGDLQVTSAGHCTTTAQKAVGKDNFALIPGTNGIEERMSWMEKCVASG 373
 Db 298 QEALWRLGSLGNLHTTADHCCFCAEQKAMGRDFFKIPNGTAGIEDRMALLWDAGVNSG 357
 QY 374 KMDNEFVAVTSTNAKIFNFYPRKGRVAVGSDADLVINPKATKIISAKTHNLNVEYNI 433
 Db 358 RLMSHEFVLTSTNTAKIFNLPKRAIRVAGADADLVLDPOGSRSLSAATHQRYDFNI 417
 QY 434 FEGVCEGAPAVVISQGRVALEDGKMFVTPGAGRFVPRKTFDFVYKRIKAR 485
 Db 418 FEGRTVVGIPSHITISQGLLWAGDLRAEPGAGRYVERPAYPS-VVEVLGRR 468

RESULT 10
 Q9FMP3
 ID Q9FMP3 PRELIMINARY; PRT; 531 AA.
 AC Q9FMP3;
 DT 01-MAR-2001 (Tremblrel. 16, Created)
 DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
 DE DIHYDROPYRIMIDINASE.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
 OC Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=COLUMBIA.
 RX MEDLINE-98162728; PubMed-9501997;
 RA Nakamura Y., Sato S., Kaneko T., Kotani H., Asamizu E., Miyajima N.,
 RA Tabata S.;
 RT "Structural analysis of Arabidopsis thaliana chromosome 5. III.

RT Sequence features of the regions of 1,191,918 bp covered by seventeen
 RT physically assigned pl clones*;
 RL DNA Res. 4:401-414(1997).
 DR EMBL; AB007727; BAB10038.1; --
 SQ SEQUENCE 531 AA; 57991 MW; F6DE523D73027789 CRC64;

Query Match 34.9%; Score 1043; DB 10; Length 531;
 Best Local Similarity 44.4%; Pred. No. 3.3e-67;
 Matches 209; Conservative 84; Mismatches 166; Indels 12; Gaps 5;
 QY 12 ITSDRLIRGGRIVNDQSFYADVHVEDGLIKOIGENLIVPGGIHTIDAGLMLVPGGVY 71
 Db 42 VSTRILKGGTVVNAHQELADYVENGIIIVAVQIKGVDEVTLDAATGKRVMPGGID 101
 QY 72 VHTRLQMPVLGMPADDFCOGTTKAALAGCTTMLDHFVDPDTGVSLLAAYEQWRERADSNA 131
 Db 102 PHTHLAMEFGTETIDDFSGQAALAGCTTMMHIDFVIPVNG-NLVAGFEAY-ENKSRRES 159
 QY 132 CDDSLHVDITRWHESIKEELALYKEGVNSFLVFMAYKDCQSDSQSOMYEIIFIRDL 191
 Db 160 CMDYGFHMAITKDEGVSRDMEMLVKEGINSKFFLAYKGLMVTDDLLLEGLKCKSL 219
 QY 192 GALAQVHAENGDIIVEEOKRLELGTGPEGHVLSHPVEAEAVYRAVTTIAKQANCPY 251
 Db 220 GALAVVHAENGDAVFEQGRMIELGITGPEGHALSRPVPVLEGEATARAIRLAREINTPLY 279
 QY 252 VTKVNSGAADIAQAKRGVVVFGEPITASLTGDSHYWKNWAKAAAFVTSPPVNDP 311
 Db 280 VHVNSVDAMDEIAKARKSGKGVIGEPVWSGLIILDDHWDLPDFTIASKYVNSPIRP- 337
 QY 312 TTADH--LACLSSGDLQVTSAGHCTTTAQKAVGKDNFALIPGTNGIEERMSWMEK 368
 Db 338 --VHGKALQDALSTGLQLVGDHCTFNSTQKALGDLDFRRIPNGVNGLEERHMLIWD 395
 QY 369 CVASGKMDNEFVAVTSTNAKIFNFYPRKGRVAVGSDADLVINPKATKIISAKTHNLN 428
 Db 396 MVESGQLSATDYVRITSTECARIFNIYPRKAILAGSDADIILNPNSSYEISSKSHSR 455
 QY 429 VEYNIEGVEGRCAPAVVISQGRVALEDGKMFVTPGAGRFVPRKTFDFVY 479
 Db 456 SDTNVYEGRRGKGVETIAGGRIVWNEELKVVPRSGKYI---EMPPFSY 503

RESULT 11
 Q9VND9
 ID Q9VND9 PRELIMINARY; PRT; 332 AA.
 AC Q9VND9;
 DT 01-MAY-2000 (Tremblrel. 13, Created)
 DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
 DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
 DE CRMP PROTEIN.
 GN CRMP OR BCDA:HL02693 OR CGL411.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE-20196006; PubMed-10731132;
 RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brockstein P., Brotter P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,

RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Fabios B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrera S., Fleischmann W.,
 RA Fostler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Hostin D., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Jallali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Jallali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.C., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.Y., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.M., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spler E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weltsenbach J.,
 RA Williams S.M., Woodage T., Worley J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RA "The genome sequence of *Drosophila melanogaster*.";
 RT Science 287:2185-2195(2000).
 DR EMBL; AE003602; AAF52003.2;
 DR FlyBase; FBgn0023023; CRM.
 DR InterPro; IPR002195;
 DR Pfam; PF00744; Dihydroorotase; 1.
 SQ SEQUENCE 332 AA; 36453 MW; 97BF4D2CFF3F601 CRC64;

Query Match 28.08; Score 834.5; DB 5; Length 332;
 Best Local Similarity 51.4%; Pred. No. 1.9e-52;
 Matches 166; Conservative 50; Mismatches 100; Indels 7; Gaps 3;
 QY 7 KSP---RTSRLLRGRVNDQSFYADVHVHEDGLIKQIG--ENLIVPGGIHTIDAH 61
 DB 9 KVPVHLQSAQNRVYKNGEIVNHDKSKADYVIEDGLIKFVGPSSSETIPGGVRTIDAS 68
 QY 62 GLMVLPGGVVDVHTRQLMPVLGTPADDFCOGKTAALAGTMTILDHVPDPDGVSLAAEY 121
 DB 69 GLMIIPGGIDPHTHQLPFGGAVAVDDFYHGKAAVAGGTMTIDFVLPNKHESMIEAYD 128
 QY 122 QWRERADSAACDYSLVHVDITRWHSIKERLEALVKRGVNSFLVFMAYKDRCCQSDSQM 181
 DB 129 KRSWADPKVCCDYGHLVGITWWSKSVSEIGILCKELGVNSFKTFMAYKGLYLQUNSD 188
 QY 182 YEIFSIIRDLGALAOVHAENGDIIVEEQRRLLELGTITGPEGHVLSHPPEVEAEVYRAVT 241
 DB 189 LDVFERIRHLNGVAMVHAENGDIITAKTORLLAEGINGPEGHLSRPEVEAEVYRAVCV 248
 QY 242 IAKQANCPVTVKVMKSGAADAQAQR--RGVVFSGEPITASLGTGDSHYWKNWAKAA 299
 DB 249 LAHQADCPVTVVHMVSKSAGIELARAHRYRGYINGETLAAALGTDATCCQHLGFDAAE 308
 QY 300 AFTVSPVNPDPPTADHLTCLLS 322
 DB 309 AHVLSPLPLPKTTPPEFLMKLLA 331
 RESULT 12
 ID O69809 PRELIMINARY; PRT; 467 AA.
 AC O69809;
 DT 01-AUG-1998 (TrEMBLrel. 07, Created)
 DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
 DE D-HYDANTOINASE (EC 3.5.2.2) (DIHYDROPYRIMIDINASE) (DHPASE).
 GN SCIA6.04.
 OS Streptomyces coelicolor.

OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
 OX NCBI_TaxID=1902;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2);
 RC Oliver K., Harris D.;
 RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2);
 RC Parkhill J., Barrell B.G., Rajandream M.A.;
 RA Redenbach M., Kieser H.M., Denapante D., Eichner A., Cullum J.,
 RA Kinashi H., Hopwood D.A.;
 RT "A set of ordered cosmids and a detailed genetic and physical map for
 the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
 RL Mol. Microbiol. 21:77-96(1996).
 CC -1- CATALYTIC ACTIVITY: 5,6-DIHYDROURACIL + H(2)O = 3-
 UREIDOPROPIONATE.
 CC -1- SIMILARITY: BELONGS TO THE DEHYDROPYRIMIDINASE FAMILY.
 DR EMBL; AL023496; CAA18902.1;
 DR InterPro; IPR002195;
 DR Pfam; PF00744; Dihydroorotase; 1.
 KW Hydrolyase.
 SQ SEQUENCE 467 AA; 50748 MW; CA4720374DEB0F0B CRC64;

Query Match 23.2%; Score 692.5; DB 2; Length 467;
 Best Local Similarity 33.8%; Pred. No. 5.9e-42;
 Matches 153; Conservative 84; Mismatches 192; Indels 43; Gaps 10;
 QY 14 SDRLLRGRVNDQSFYADVHVHEDGLIKOI-----GENLIVPGGIHTIDAH 61
 DB 2 SSRTVIRGLVITASDEIHDVIEDGRVLAALATGTPAAEAFTAEV-----IDAS 53
 QY 62 GLMVLPGGVVDVHTRQLMPVLGTPADDFCOGKTAALAGTMTILDHVPDPDGVSLAAEY 121
 DB 54 GKTVIFGGVDGHTHEMPPGGTYAADTFETGTRAAAWGGTTTIVDFAIQSVGHSRLREG 113
 QY 122 QWRERADSAACDYSLVHVDITRWHSIKERLEALVKRGVNSFLVFMAYKDRCCQSDSQM 181
 DB 114 AWHAKAEGCAIDYGFHMTVSDVNOETLKEMDILV--EEGVTSFKQFMAYPGVYSDGQI 172
 QY 182 YEIFSIIRDLGALAOVHAENGDIIVEEQRRLLELGTITGPEGHVLSHPPEVEAEVYRAVT 241
 DB 173 LFAMQRAENGGLIMHMAENGTAIDVLVEQALARGTDPFRHGEVRKALLEAEATHRAIR 232
 QY 242 IAKQANCPVTVKVMKSGAADAQAQRGVVVFGE--PITASLGTGDSHYWKNWAKAA--- 296
 DB 233 LAQVAGAPLVVHVVSATEAVAELTRARDEGLPVFGTCQYLFLESTD-----NLAEPD 285
 QY 297 -KAAAFVTSPPVNPDPPTADHLTCL---LSSGDLQVTSAGHCTFT--TAQAVKGNFALI 351
 DB 286 FEGAKYVCSPLRP-----KEHQALNRLGLTNDLVVSTDHCPFCFSGKELGRGDSRI 341
 QY 352 PEGTNGIEERMSVWKECVASGKMDENEFVAVTSTNAKIFNFYPRKGRVAVGSDADLVI 411
 DB 342 PNGMPGVENRMDLL--HQAVVEGHIGRRRWEIACATPAMFGLYPKKGTITACADADIV 400
 QY 412 WNPATKTIISAKTHNLNVEYNIFGVECRGAPAVVISQGRVALEDGKMFVTPGAGRFVR 471
 DB 401 YDPAEQVISAETHMNVDSIAYESGRRTIGRVETVLSRGEPVVTREYTGKRGHGYATPR 460
 QY 472 KT 473
 DB 461 AT 462

```

RESULT 13
Q9P903
ID Q9P903 PRELIMINARY; PRT; 542 AA.
AC Q9P903;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE DIHYDROXYRIMIDINASE (EC 3.5.2.2).
GN PYD2.
OS Saccharomyces kluyveri (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4934;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NRRL Y-12651;
RX MEDLINE=20123992; PubMed=10656811;
RA Gofkovic Z., Jahne K., Schnackerz K.D., Piskur J.;
RT "PYD2 encodes 5,6-dihydroxyrimidine amidohydrolase, which participates
RT in a novel fungal catabolic pathway.";
RL J. Mol. Biol. 295:1073-1087(2000).
DR EMBL: AF156967; AAF69237.1; -
DR InterPro; IPR002195; -
DR Pfam; PF00744; Dihydroorotase; 1.
KW Hydrolase.
SQ SEQUENCE 542 AA; 60240 MW; E8758452CB86E98 CRC64;

Query Match 17.5%; Score 523.5; DB 3; Length 542;
Best Local Similarity 29.1%; Pred. No. 1.2e-29;
Matches 153; Conservative 92; Mismatches 207; Indels 73; Gaps 18;

QY 17 LLIRGRIVNDQSFYADVHVEDGLIKOIGENLIVPGIHTIDAGHLMVLPGGVDVHTRL 76
Dy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 6 LIKNGIICTASDIYAAFIWNGKVLIASIDPSLGSSEVIDAEGAFITPGGIDAHVH 65
Dy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 77 QMP--VLGMPADDFCOGTTKAALAGGTTMILDHVPDT---GVSLAAAYEOW--RERADS 129
Dy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 66 DEPLKLLG-DVVDPTMEHATRSVAGGTTTVAFSTQDVSKKPSALAESVKLDVDEYSEQ 124
Dy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 130 AACDYSILHVDI-----TRWHESIKKEALVKEGVNSFLVFMAYKDRQCSDSQMY 182
Dy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 125 TLXCDYGLHLILFQIEKPSVEARELDVQQAAYNDYGVSSKMFMTYPG-LQISD---Y 180
Dy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 183 EIPSI---RDGLAQVHAENGDIIVEEQKRLLELIGITGPEGHVLSHPPEVEAEAVYRA 239
Dy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 181 DIMSAMYATKNGFTTMLHAENGDMVKWMEALEEOGLDADYHGVSRPSIVEGEATNRA 240
Dy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 240 VTIAQANCLPYTVKVMKGAADAIQAQKRGVVVF-----GPEPI 279
Dy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 280 TASLQTDGSHYKSNKNAK-----AAAFVTSPPVNPDPPTADHLTCLSSGDLQVYSAH 333
Dy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 301 SYGVGIDLSSESPTNPDDRFISGKYICSPPIRPEGTQKS-IWGMNNFTTIVGSDH 359
Dy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 334 CTF-----TTAQK--AVGKDN-----FALIDEGTNGIEERMSMWKCVASGKMDNE-NE 379
Dy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 360 CSYNYEKTSTASKARAFDPENKNNGEFRIYIPNGLPGVCTRMPLLYDYGLRGNLTSMK 419
Dy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 380 FVAVTSTNAKIFNFPKRGVAVG--SDADLVVWNP-----KATKIIISAKTHNLNVE 430
Dy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 420 LVEIOCTNPAKYGVYKQKSLIPGVSDADLVVWNPDSKEYSKPKLITNKLMHNCD 479
Dy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 431 YNIFGVECRGAPAVVISQGRVALEDGKMFVTPGAGRVPR-KTF 474
Dy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 480 YTFEGIEIKNPRTYIVKGIKIVYKEGELIKENADGKYLKRGKSF 524
Dy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 14
Q9F465
ID Q9F465 PRELIMINARY; PRT; 458 AA.
AC Q9F465;

```

```

DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE L-HYDANTONINASE HYUH.
GN HYUH.
OS Arthrobacter aureus.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Micrococcales; Micrococcaceae; Arthrobacter.
OX NCBI_TaxID=43663;
RN [1]
RP SEQUENCE OF 451-458 FROM N.A.
RC STRAIN=DSM3747;
RX MEDLINE=99210756; PubMed=10194852;
RA Wilms B., Wiese A., Syltack C., Mattes R., Altenbuchner J.,
RA Pietzsch M.;
RT "Cloning, nucleotide sequence and expression of a new L-N-carbamoylase
RT gene from Arthrobacter aureus DSM 3747 in E. coli.";
RL J. Biotechnol. 68:101-113(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=DSM3747;
RX MEDLINE=20403588; PubMed=10949312;
RA Wiese A., Pietzsch M., Syltack C., Mattes R., Altenbuchner J.;
RT "Hydantoin racemase from Arthrobacter aureus DSM 3747: heterologous
RT expression, purification and characterization.";
RL J. Biotechnol. 80:217-230(2000).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=DSM3747;
RX Wiese A.H., Altenbuchner J.;
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF146701; AAG02130.1; -
SQ SEQUENCE 458 AA; 49659 MW; 9321FD948B264994 CRC64;

Query Match 13.5%; Score 402.5; DB 2; Length 458;
Best Local Similarity 28.2%; Pred. No. 5.2e-21;
Matches 134; Conservative 85; Mismatches 208; Indels 49; Gaps 13;

QY 17 LLIRGRIVNDQSFYADVHVEDGLIKOIGENLIVPGIHTIDAGHLMVLPGGVDVHTRL 76
Dy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 4 VIVKNCRVSSDGTIEADILVKDGKVAISADTRDVEASRTIDAGGRFVMPGVDEHVH- 62

QY 77 QMPVLGMPADDFCO---GTRAAALAGGTTMILDH--VFPDTGVSLLAAYEQWRADSA 131
Dy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 63 ---IIDMLKRYGRFELDSAAVGGITTIEMPTFPPT--TTDAFLKKKQAGORL 117

QY 132 CDYSLH-----VDITRWHSIKKEALVKEGVNSFLVFMAYK----DRCOCSD 178
Dy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 118 KVDFALYGGVPGNLPETKMHDA-----GAVGFKSMMAASVPMFMD--AVSD 163

QY 179 SOMYEIETSIIRDLGALQVHAENGDIIVEEQKRLLELIGITGPEGHVLSHPPEVEAEAVYR 238
Dy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 164 GELPEIQTAEACGSVIVHAENETIIQALQKIKAAAGKDMAAYEASQPVFQENAIOR 223

QY 239 AVTIAQANCLPYTVKVMKGAADAIQAQKRGVVVFGE--PITASLIGTDSHYSKNA 296
Dy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 224 ALLQKEAGRLIVLHVSNDPVGVELIHQAQSEGQVHCEGPOYLNITTTDAERIGP-YM 282

QY 297 KAAAFVTSPPVNPDPPTADHLTCLSSGDLQVYSAHCTFTTAQKAVGKDNFALIEGTN 356
Dy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 283 KVAPPVRSAMNV-----RLWQLENGVITDITLGSDDHGHVPEDKEPQKWKVAKAGNAL 336

QY 357 GIERMSMWKCVASGKMDENEFVAVTSTNAKIFNFPKRGVAVGSDADLVVWNPKA 416
Dy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 337 GLETSPLMMLTNGVNGKRLSLERLVEVMECKPAKLGIVPQKGTQVQSDADLLIILDDI 396

QY 417 -TKIISAKTHNLNVEINIFEGVECRGAPAVVISQGRVALEDGKMFVTPGAGRVPR 471
Dy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 397 DTKVDASQFRSLH-KYSPDPGMPVTGAPVLTWVRGTVAEQGVELVVEQGFQGVTR 451

RESULT 15

```

us-09-367-496-8.std.rspt

•Tue Jul 31 13:08:21 2001

Q9KAH8 PRELIMINARY; PRT: 438 AA.

ID Q9KAH8; 2000 (Tremblrel. 15, Created)

AC Q9KAH8; 2000 (Tremblrel. 15, Last sequence update)

DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)

DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)

DE ALLANTOININASE.

GN BH2309.

OS Bacillus halodurans.

OC Bacteria; Firmicutes; Bacillus/Clostridium group;

OC Bacillus/Staphylococcus group; Bacillus.

OX NCBI_TaxID=86665;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=C-125 / JCM 9153;

RA Takami H., Nakasone K., Takaki Y.;

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

EMBL: AP001515; BAB06028.1; -

DR EMBL: IPR002195; -

DR InterPro: IPR002195; -

DR Pfam: PF00744; Dihydroorotase; 1.

DR PROSITE: PS00482; DIHYDROOROTASE.1; UNKNOWN.1.

SQ SEQUENCE 438 AA; 48542 MW; 47A24DD790AF947F CRC64;

Query Match 10.6%; Score 317; DB 2; Length 438;

Best Local Similarity 24.8%; Pred. No. 7.4e-15;

Matches 117; Conservative 80; Mismatches 204; Indels 70; Gaps 16;

QY 17 LLIRGGRIIVNDQSFYADVHVEDGLIKQIGENLIVPGGIHT-----IDAHGLMVLPGGVD 71

DB 6 LIIRSTVVTETTYRADVAIRGIVSAITE-----PGSISSDDGPAIDGTGLHLFGMYD 61

QY 72 VTRLQMPVLGMPADDFCQCTKAALAGTMTILDHVFDPDGTGVSLL-AAEOWNRERADSA 130

DB 62 VHVHNEP--GRTEWEGFASGSKSLAAGGVTYTFDPLNSNPPTITREELDKKROLANEK 119

QY 131 ACCDYSLVHVDITRWHESTKEELEAL--VKEKGVNSFLVFMAYKDRCCSDSQMYEIFSII 188

DB 120 SLVDYRF-----WGGLVPGNIDHLQDLHDGCVIGFKAFMS---ECGTDDDFQFSDHETLL 170

QY 189 RD-----LGALAQVHAENGDIIVE-----EOKRLLELGIITGEGHVLSHPEEVEAEA 235

DB 171 KGMKKTAAALGSLAVHAESNMVNALTTIAEEQRLTV-----KDYSEARPIVSELEA 223

QY 236 VTRAVTIAQANCPLYVTVMKSGAADAIAQAKRGV-----VVFGEPTIASLGT 285

DB 224 VERILFAQLTCCPIHCHVSRKVLKRIKQAKGEGVNVSVETCPHYLLFSLDEFAEIG- 282

QY 286 DGSYHWSKNWAKAAAEVTPPVNPDPTTADHLTCLISSDQLQVTGSAHCTFTTAQKAVGK 345

DB 283 -----YLAK-----CAPPLRERQEVED-LWDGLMAGEIDLISSDHSP-SLPOMKTGK 327

QY 346 DNFAIPEGTNGTEERMSVMWEKCVASGKMDENEFVATVTNAAKIFNFYPRKGRVAVGS 405

DB 328 TIFE-VWGGIAGGCONTLAVMLTEGYHKRMPLTQIVQLLSTEPAKRFRGLYPQKGTIQVGA 386

QY 406 DADLVINWPKATKIISAKTNHNLNVEYNIEFGEVCEGRCAPAVVISQGRVALED 456

DB 387 EASFTLIDLNESTYLNASDLRYRHPISPYVQRFGRGKVKHHTICQGHVYQD 437

Search completed: July 30, 2001, 11:42:00

Job time: 229 sec

OM protein - protein search, using sw model

Run on: July 30, 2001, 11:40:01 ; Search time 10.27 Seconds
(without alignments)
1927.900 Million cell updates/sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

Scoring table: BLOSUM62

searched: 93435 seqs, 34255486 residues

93435 chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

[illegible]

Post-processing: Minimum Match 0%
Minimum Match 100%

Maximum Match 100%
 Missing first 45 summaries

Database : SwissProt_39:*

SUMMARIES

Result No.	Score	Query		ID	Description
		Match	Length		
1	2972	99.6	572	DPY4_HUMAN	O14531 homo sapien
2	2818	94.4	572	DPY4_MOUSE	O35098 mus musculus
3	2789	93.4	564	DPY4_RAT	O32951 rattus norv
4	2362	79.1	572	DPY2_CHICK	O90635 gallus gall
5	2347	78.6	572	DPV2_BOVIN	O02675 bos taurus
6	2343	78.5	572	DPY2_RAT	P47942 rattus norv
7	2339	78.4	572	DPY2_HUMAN	O16555 homo sapien
8	2325	77.9	572	DPY2_MOUSE	O08553 mus musculu
9	2187	73.3	570	DPY3_MOUSE	O32198 mus musculu
10	2186	73.2	570	DPY3_HUMAN	O14195 homo sapien
11	2175.5	72.8	571	DPY3_XENLA	O13022 xenopus lae
12	2152	72.3	572	DPV1_MOUSE	P97427 mus musculu
13	2152	72.1	572	DPV1_RAT	O62950 rattus norv
14	2151	72.1	572	DPY1_HUMAN	O14194 homo sapien
15	1554	52.1	519	DPY5_HUMAN	O14117 homo sapien
16	1548	51.9	519	DPY5_RAT	O63150 rattus norv
17	1451	48.6	358	DPY3_RAT	O29552 rattus norv
18	1245	41.7	520	DPV1_CAEEL	O18677 caenorhabdi
19	1122	37.6	544	DPY2_CAEEL	O21773 caenorhabdi
20	1040	34.8	495	HYDA_PSEPU	O95699 pseudomonas
21	837	28.0	471	HYDA_BACST	O45515 bacillus st
22	781.5	26.2	854	UN33_CAEEL	O01630 caenorhabdi
23	691.5	23.2	457	YGE2_ECOLI	O44184 agrobacteri
24	672	22.5	465	HYDL_ARTAU	O64806 escherichia
25	406.5	13.6	458	ALN_ARTAU	P81006 arthrobacte
26	325.5	10.9	448	ALN_ECOLI	O9r76 deinococcus
27	311.5	10.4	453	ALN_ECOLI	P77671 escherichia
28	296.5	9.9	430	PYRC_MYCTU	P71809 mycobacteri
29	295.5	9.9	446	ALN_BACSU	O32137 bacillus su
30	295.5	9.9	454	PYRC_METTH	O27199 methanobact
31	257.5	8.6	445	ALN_STRCO	O9rku5 streptomyce
32	241.5	8.1	422	PYRC_AQUAE	O66990 aquifex aeo
33	229.5	7.7	2242	PYR1_SQAC	O91437 squalus sca

Db 481 RIKARSLAEIRGVRGLYDGPVCEVSTPKVTPTASSAKTSPAKQAPPVRLNLSQSGFS 540
Qy 541 LSGSQADDDHARRTAQKIMAPPGRSNTISL 571
Db 541 LSGAQIDNDIPRRTTQRIVAPPGRANITSL 571
RESULT 5
DPY2_BOVIN STANDARD; PRT; 572 AA.
AC 002675;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE DIHYDROXYRIMIDINASE RELATED PROTEIN-2 (DRP-2) (NEURAL SPECIFIC PROTEIN
NSP60).
GN DPYSL2.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RA Kamata T.K.;
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE DEHYDROXYRIMIDINASE FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
DR EMBL; U83278; AAB80618.1;
DR HSP; P18316; IKRC.
DR InterPro; IPR002195;
DR Pfam; PF00744; Dihydroorotase; 1.
SQ SEQUENCE 572 AA; 62277 MW; 343507ACB9D91BDE CRC64;

Query Match 78.6%; Score 2347; DB 1; Length 572;
Best Local Similarity 75.7%; Pred. No. 3.2e-162;
Matches 432; Conservative 72; Mismatches 67; Indels 0; Gaps 0;
Qy 1 MSFOGKKSIPRITSRLIRGRVNDQSFYADVHVHEDGLIKQIGENLIVPGGIHTIDA 60
Db 1 MSYQGNKIPRITSRLIRGRVNDQSFYADVHVHEDGLIKQIGENLIVPGGIHTIDA 60
Qy 61 HGLMVLPGVDVHTRLQMPVLCMTADDFOCGTKAALAGGTTMLDHFVPTDGYSLAAY 120
Db 61 HSRWIPGIDVHTRFQMPDQGMTSADDFQCGTKAALAGGTTMLDHFVPTDGYSLAAY 120
Qy 121 EWRERADSAACCDYSLHVDITRWHEISKELEALVKEGWSFLVPMAYKDRQCSDSQ 180
Db 121 DQREWADSKCCDYSLHVDITRWHEISKELEALVKEGWSFLVPMAYKDRQCSDSQ 180
Qy 181 MYEFTSIRDLGALQAQVAENGDIIVEEQKRLLELGTGPEGHVLSHPEVEAEVYRAV 240
Db 181 IYELSVIRDIGALQAQVAENGDIIVEEQKRLLELGTGPEGHVLSHPEVEAEVYRAV 240
Qy 241 TIAQANCPVTVTKVMSKGAATAQAKRGVVVFGPEPTASLGTGDSHWYKSNWAKAAA 300
Db 241 TIANQNCPLVTVTKVMSKGAATAQAKRGVVVFGPEPTASLGTGDSHWYKSNWAKAAA 300
Qy 301 FVTSPPVNPDPPTADHLTCLSSGDLQVTSACHTFTTAQKAVGKDNFALIPGTNGIEE 360
Db 301 FVTSPPVNPDPPTADHLTCLSSGDLQVTSACHTFTTAQKAVGKDNFALIPGTNGIEE 360

Qy 361 RMSWWEKCVASKMDENEFVAVTSTNAKIFNFPKRGVAVGSDADLVWNPATKII 420
Db 361 RMSVWIDKAVTGKMDENEFVAVTSTNAKIFNFPKRGVAVGSDADLVWNPATKII 420
Qy 421 SAKTHNLNVEYNLEFEGVECRGAPAVVISOGRVALEDEGKMFVTPGAGRFVPRKTFDFEYK 480
Db 421 SAKTHNLNVEYNLEFEGVECRGAPAVVISOGRVALEDEGKMFVTPGAGRFVPRKTFDFEYK 480
Qy 481 RIKARNRLAEIHGVRGLYDGPVCEVSTPKVTPTASSAKTSPAKQAPPVRLNLSQSGFS 540
Db 481 RIKARSLAEIRGVRGLYDGPVCEVSTPKVTPTASSAKTSPAKQAPPVRLNLSQSGFS 540
Qy 541 LSGSQADDDHARRTAQKIMAPPGRSNTISL 571
Db 541 LSGAQIDNDIPRRTTQRIVAPPGRANITSL 571
RESULT 6
DPY2_RAT STANDARD; PRT; 572 AA.
AC P47942;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE DIHYDROXYRIMIDINASE RELATED PROTEIN-2 (DRP-2) (TURNED ON AFTER
DIVISION, 64 KDA PROTEIN) (TOAD-64) (COLLAPSED RESPONSE MEDIATOR
PROTEIN 2) (CRMP-2).
GN DPYSL2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.; AND SEQUENCE OF 136-142; 402-418; 441-450 & 499-511.
RC STRAIN-SPRAGUE-DAWLEY; TISSUE=Brain;
RX MEDLINE=96033765; PubMed=7472434;
RA Minton J.E., Fryer H.J.L., Geschwind D.H., Hockfield S.;
RT "TOAD-64, a gene expressed early in neuronal differentiation in the
RT rat, is related to unc-33, a C. elegans gene involved in axon
RL J. Neurosci. 15:6757-6766(1995).
CC -1- FUNCTION: MAY HAVE A ROLE IN AXON ELABORATION.
CC -1- SUBCELLULAR LOCATION: TIGHTLY, BUT NONCOVALENTLY, ASSOCIATED WITH
CC MEMBRANES.
CC -1- DEVELOPMENTAL STAGE: EXPRESSED IMMEDIATELY AFTER NEURONAL BIRTH
CC AND IS DRAMATICALLY DOWNREGULATED IN THE ADULT.
CC -1- SIMILARITY: BELONGS TO THE DEHYDROXYRIMIDINASE FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
DR EMBL; Z46882; CAA86981.1;
DR InterPro; IPR002195;
DR Pfam; PF00744; Dihydroorotase; 1.
SQ SEQUENCE 572 AA; 62277 MW; C031F3BC038AA737 CRC64;

Query Match 78.5%; Score 2343; DB 1; Length 572;
Best Local Similarity 75.8%; Pred. No. 6.2e-162;
Matches 433; Conservative 70; Mismatches 68; Indels 0; Gaps 0;
Qy 1 MSFOGKKSIPRITSRLIRGRVNDQSFYADVHVHEDGLIKQIGENLIVPGGIHTIDA 60
Db 1 MSYQGNKIPRITSRLIRGRVNDQSFYADVHVHEDGLIKQIGENLIVPGGIHTIDA 60
Qy 61 HGLMVLPGVDVHTRLQMPVLCMTADDFOCGTKAALAGGTTMLDHFVPTDGYSLAAY 120
Db 61 HSRWIPGIDVHTRFQMPDQGMTSADDFQCGTKAALAGGTTMLDHFVPTDGYSLAAY 120

"Characterization of the human dihydropyrimidinase-related protein 2

(DRP-2) gene."

RT DNA Res. 6:291-297(1999).

RL -!- SUBCELLULAR LOCATION: CYTOPLASMIC (POTENTIAL).

CC -!- TISSUE SPECIFICITY: UBIQUITOUS.

CC -!- SIMILARITY: BELONGS TO THE DEHYDROPYRIMIDINASE FAMILY.

CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).

CC

DR EMBL: U17279; AAA93202.1; -

DR EMBL: D78013; BAA11191.1; -

DR EMBL: U97105; AAC05793.1; -

DR EMBL: A020777; BAA86991.1; JOINED.

DR EMBL: A020764; BAA86991.1; JOINED.

DR EMBL: A020765; BAA86991.1; JOINED.

DR EMBL: A020766; BAA86991.1; JOINED.

DR EMBL: A020767; BAA86991.1; JOINED.

DR EMBL: A020768; BAA86991.1; JOINED.

DR EMBL: A020769; BAA86991.1; JOINED.

DR EMBL: A020770; BAA86991.1; JOINED.

DR EMBL: A020771; BAA86991.1; JOINED.

DR EMBL: A020772; BAA86991.1; JOINED.

DR EMBL: A020773; BAA86991.1; JOINED.

DR EMBL: A020774; BAA86991.1; JOINED.

DR EMBL: A020775; BAA86991.1; JOINED.

DR EMBL: A020776; BAA86991.1; JOINED.

DR HSP; P18316; ICR.

DR MIM: 602463; -

DR InterPro: IPR002195; -

DR Pfam: PF00744; Dihydroorotase; 1.

DR

SQ SEQUENCE 572 AA; 62293 MW; 5CDB6CF7F5C308AD CRC64;

Query Match 78.4% Score 2339; DB 1; Length 572;

Best Local Similarity 75.5% Pred. No. 1.2e-161;

Matches 431; Conservative 72; Mismatches 68; Indels 0; Gaps 0;

Qy 1 MSFQCKSIPRITSDRLIRGRVINDQSFYADVHVEDGLIKQIGENLVPGGIHTIDA 60

Db 1 MSYQCKNIPRITSDRLIRGRVINDQSFYADYIMEDGLIKQIGENLVPGGVKTIEA 60

Qy 61 HGLMVLPGVDVHTLQMPVLTGMPADDFCQGTAAALAGGTTMLDHFVDPDTGVSLLAAY 120

Db 61 HSRMVPVGGIDVHTRFQMPDQGTADDFFQGTAAALAGGTTMLDHFVDPDTGVSLLAAY 120

Qy 121 EQWRERADSAACDYSILHVDITRWHESIKEELEALVKEGVNSFLVFMAYKDRCCQSDQ 180

Db 121 DQWRWADSKSCDYSILHVDISEWHKIQEEMALVKDHGVNSFLVFMAYKDRFQLTDCQ 180

Qy 181 MYEIFSIRDLGALAQVHAENGDIIVEEQKRLLELGLITGPEGHVLSHPPEVEAEAVRAY 240

Db 181 IYEVLSVIRDIGALAQVHAENGDIIVEEQKRLLELGLITGPEGHVLSHPPEVEAEAVRAY 240

Qy 241 TIAQANCPLYTVKVMKGAADAIAQAKRRGVVVFGEPIITASLGTDGSHYKSNWAKAAA 300

Db 241 TIANQNCPLIYTVKVMKSSAEVIAQARKGVVYGEPIITASLGTDGSHYKSNWAKAAA 300

Qy 301 FVTSPPVNPDPPTADHLTCLSSGDLQVTSACHTFTTAQKAVGKDNFALIPGTINGIEE 360

Db 301 FVTSPPVNPDPPTADHLTCLSSGDLQVTSACHTFTTAQKAVGKDNFALIPGTINGIEE 360

Qy 361 RSMNWEKCVASGKMDENEFVAVTSTNAKIFNFKRGVAVGSDADLVINPKATKII 420

Db 361 RSMVINDKAVVTGKMDENOFVAVTSTNAKIFNFKRGVAVGSDADLVINPDPSVKTI 420

Qy 421 SAKTHNLNVEYNIFEGVECRGAPAVVISQGRVALEDKGMFVTPGAGRFVPRKTFDFVYK 480

Qy 121 EQWRERADSAACDYSILHVDITRWHESIKEELEALVKEGVNSFLVFMAYKDRCCQSDQ 180

Db 121 DQWRWADSKSCDYSILHVDITRWHESIKEELEALVKEGVNSFLVFMAYKDRFQLTDCQ 180

Qy 181 MYEIFSIRDLGALAQVHAENGDIIVEEQKRLLELGLITGPEGHVLSHPPEVEAEAVRAY 240

Db 181 IYEVLSVIRDIGALAQVHAENGDIIVEEQKRLLELGLITGPEGHVLSHPPEVEAEAVRAY 240

Qy 241 TIAQANCPLYTVKVMKGAADAIAQAKRRGVVVFGEPIITASLGTDGSHYKSNWAKAAA 300

Db 241 TIANQNCPLIYTVKVMKSSAEVIAQARKGVVYGEPIITASLGTDGSHYKSNWAKAAA 300

Qy 301 FVTSPPVNPDPPTADHLTCLSSGDLQVTSACHTFTTAQKAVGKDNFALIPGTINGIEE 360

Db 301 FVTSPPVNPDPPTADHLTCLSSGDLQVTSACHTFTTAQKAVGKDNFALIPGTINGIEE 360

Qy 361 RSMNWEKCVASGKMDENEFVAVTSTNAKIFNFKRGVAVGSDADLVINPKATKII 420

Db 361 RSMVINDKAVVTGKMDENOFVAVTSTNAKIFNFKRGVAVGSDADLVINPDPSVKTI 420

Qy 421 SAKTHNLNVEYNIFEGVECRGAPAVVISQGRVALEDKGMFVTPGAGRFVPRKTFDFVYK 480

Db 421 SAKTHNSALEYNIFEGVECRGAPAVVISQGRVALEDKGMFVTPGAGRFVPRKTFDFVYK 480

Qy 481 RIKARSLAETHGVPRGLYDGPVHEVMVPAKPGSGAPARASCPGKISVPPVRNLHQSGFS 540

Db 481 RIKARSLAETHGVPRGLYDGPVHEVMVPAKPGSGAPARASCPGKISVPPVRNLHQSGFS 540

Qy 541 LSGSQADHIAARTTAQKIMAPPGGRSNTISL 571

Db 541 LSGAQIDDNIPRRTTORIVAPPGGGRANITSL 571

RESULT 7

DPY2_HUMAN STANDARD; PRT; 572 AA.

AC Q16555; O00424;

DT 15-JUL-1998 (Rel. 36, Last sequence update)

DT 15-JUL-1998 (Rel. 36, Last sequence update)

DT 01-OCT-2000 (Rel. 40, Last annotation update)

DE DIHYDROPYRIMIDINASE RELATED PROTEIN-2 (DRP-2) (COLLAPSE IN RESPONSE

DE MEDIATOR PROTEIN 2) (CRMP-2) (N2A3).

GN DPYSL2 OR CRMP2.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=95364923; PubMed=7637782;

RA Goshima Y., Nakamura F., Strittmatter P., Strittmatter S.M.;

RT "Collapsin-induced growth cone collapse mediated by an intracellular

RT protein related to UNC-33."

RL Nature 376:509-514(1995).

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE=97128821; PubMed=8973361;

RA Hamajima N., Matsuda K., Sakata S., Tamaki N., Sasaki M., Nonaka M.;

RT "A novel gene family defined by human dihydropyrimidinase and three

RT related proteins with differential tissue distribution."

RL Gene 180:157-163(1996).

RN [3]

RP SEQUENCE FROM N.A.

RX TISSUE=Liver;

RA Zhou J., Chen Y., Gu J.R.;

RT Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.

RN [4]

RP SEQUENCE FROM N.A.

RX MEDLINE=20039612; PubMed=10574455;

RA Kitamura K., Takayama M., Hamajima N., Nakanishi M., Sasaki M.,

RA Endo Y., Takemoto T., Kimura H., Iwaki M., Nonaka M.;

Db 421 SAKTHNSLEYNIFEGMECRGSPVWISQKIVLEDTLHVTEGSGRYIPRKPPDFVYK 480
 QY 481 RIKARNRLAEIHGVRGLYDGPVHEVMVPAKPGSGAPARASCPGKISVPPVRLNHQSGFS 540
 Db 481 RIKARSRLAEHLRGVPRGLYDGPVCEVSVTPKTVTPASSAKTSPAKQAPPVRLNHQSGFS 540
 QY 541 LSGSQADHIAARRTAQKIMAPPGGRSNTLSL 571
 Db 541 LSGAQIDNIPRTTQRIVAPPGRANITSL 571

RESULT 8
 ID DPY2_MOUSE STANDARD; PRT; 572 AA.
 AC O08553;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE DHYDROPYRIMIDINASE RELATED PROTEIN-2 (DRP-2) (ULIP 2 PROTEIN).
 GN DPYSL2 OR ULIP2 OR CRMP2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ICR; TISSUE=Brain;
 RX MEDLINE=98314456; PubMed=96523388;
 RA Byk T., Ozon S., Sobel A.;
 RT "The Ulip family phosphoproteins -- common and specific properties.";
 RL Eur. J. Biochem. 254:14-24(1998).
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (POTENTIAL).
 CC -1- SIMILARITY: BELONGS TO THE DEHYDROPYRIMIDINASE FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 DR EMBL: Y10339; CAA71370.1; -
 DR MGI: 1349763; Dpysl2.
 DR InterPro: IPR002195; -
 DR Pfam: PF00744; Dihydroorotase; 1.
 SQ SEQUENCE 572 AA; 62170 MW; 7FD4E8A242ACF62D CRC64;

Query Match 77.9%; Score 2325; DB 1; Length 572;
 Best Local Similarity 75.3%; Pred. No. 1.2e-160;
 Matches 430; Conservative 70; Mismatches 71; Indels 0; Gaps 0;

QY 1 MSFGQKKSIPRTSDRLIRGRIVNDQSFYADVHVEDGLIKOIGENLIVPGGIHTIDA 60
 Db 1 MSYQCKKNIPRTSDRLIRGRIVNDQSFYADYMEDGLIKOIGENLIVPGGVKTIEA 60
 QY 61 HGLMVLPGGVVHTRQLMPVLGMTPADFCQGTAKAALAGGTTMLDHVFDPDTGVSLLAA 120
 Db 61 HSRMVIPGGIDVHTRFQMPDQGMTSADDFEQGTAKAALAGGTTMLDHVFDPDTGVSLLAA 120
 QY 121 EWRERADSAACCDYSLHVDITRHSIKKELEALVKEGVNSFLVPMAYKDRQCSDSQ 180
 Db 121 DQREWADSKCCDYSLHVDITRHSIKKELEALVKEGVNSFLVPMAYKDRFQLTDSQ 180
 QY 181 MYEFTSIIRDGALAAQVHAENGDIIVEEQRKLELGLTGTGPEGVHLSHPERVEAEVYRAV 240
 Db 181 IYEVLSVIRDTGALAAQVHAENGDIIEAQAQRILDLGTGPEGVHLSHPERVEAEVYRSI 240
 QY 241 TIAQANCPLYTVKVMKSGAADAIAQAKRGVVVVEPTTASLGTDGSHYWSKNWAKAAA 300
 Db 241 TIANQTNCPLYTVKVMKSAEYIAQARKKGTGVYGEPTTASLGTDGSHYWSKNWAKAAA 300

Query Match 73.3%; Score 2187; DB 1; Length 570;
 Best Local Similarity 70.1%; Pred. No. 1.2e-150;
 Matches 401; Conservative 90; Mismatches 79; Indels 2; Gaps 2;

QY 1 MSFGQKKSIPRTSDRLIRGRIVNDQSFYADVHVEDGLIKOIGENLIVPGGIHTIDA 60
 Db 1 MSYQCKKNIPRTSDRLIRGRIVNDQSFYADYMEDGLIKOIGENLIVPGGVKTIEA 60
 QY 61 HGLMVLPGGVVHTRQLMPVLGMTPADFCQGTAKAALAGGTTMLDHVFDPDTGVSLLAA 120
 Db 61 NGKMWIPGGIDVHTRFQMPYKGMTTVDFFQGTAKAALAGGTTMLDHVFDPDTGVSLLAA 120

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL: D78014; BAAL1192.1; -;
DR EMBL: Y07818; CA69153.1; -;
DR MIM: 601168; -;
DR InterPro: IPR002195; -;
DR Pfam: PF00744; Dihydroorotase; 1; L -> V (IN REF. 2).
FT CONFLICT 49 49 T -> A (IN REF. 2).
FT CONFLICT 142 142 T -> A (IN REF. 2).
SQ SEQUENCE 570 AA; 61963 MW; 9D6AF86CB33AD5 CRC64;

Query Match 73.2%; Score 2186; DB 1; Length 570;
Best Local Similarity 69.9%; Pred. No. 1.4e-150; Indels 2; Gaps 2;
Matches 400; Conservative 91; Mismatches 79;

QY 1 MSQGGKSIPIRSDRLIRGRIVNDQSFYADVHVEDGLIKOIGENLIVPGGIHTDA 60
DB 1 MSQGGKSIPIRSDRLIRGRIVNDQSFYADVHVEDGLIKOIGENLIVPGGIHTDA 60
QY 61 HGLMVLPGGVYVHTRLOMPVLGMPADDFCQGTAAALAGGTTMLDRHVPDFTGVSLAAY 120
DB 61 HGLMVLPGGVYVHTRLOMPVLGMPADDFCQGTAAALAGGTTMLDRHVPDFTGVSLAAY 120
QY 61 NGKMWIPGGIDVHTFQMPYKGMTVDVDFQGTAAALAGGTTMLDRHVPDFTGVSLAAY 120
DB 61 NGKMWIPGGIDVHTFQMPYKGMTVDVDFQGTAAALAGGTTMLDRHVPDFTGVSLAAY 120
QY 121 EWRERADSAACDYSLVHVDITRWHSIEKELEALVKEGVNSFLVPMAYKDRCCQSDSQ 180
DB 121 EWRERADSAACDYSLVHVDITRWHSIEKELEALVKEGVNSFLVPMAYKDRCCQSDSQ 180
QY 121 EKREWADGKSCCDYALHVDITRWHSIEKELEALVKEGVNSFLVPMAYKDRCCQSDSQ 180
DB 121 EKREWADGKSCCDYALHVDITRWHSIEKELEALVKEGVNSFLVPMAYKDRCCQSDSQ 180
QY 181 MYEFTSIIRDLGALAAQVHAENGDIIEVEEQKRLLELGTGEGHVLSPHEVEAEAVYAV 240
DB 181 MYEFTSIIRDLGALAAQVHAENGDIIEVEEQKRLLELGTGEGHVLSPHEVEAEAVYAV 240
QY 181 LYEFTCLGELGALAAQVHAENGDIIEVEEQKRLLELGTGEGHVLSPHEVEAEAVYAV 240
DB 181 LYEFTCLGELGALAAQVHAENGDIIEVEEQKRLLELGTGEGHVLSPHEVEAEAVYAV 240
QY 241 TIAQANCPYVTVKMSKGAADATAQAKRGVGVFEGPITASLGTDGSHYKSNWAKAAA 300
DB 241 TIAQANCPYVTVKMSKGAADATAQAKRGVGVFEGPITASLGTDGSHYKSNWAKAAA 300
QY 241 TIAQANCPYVTVKMSKGAADATAQAKRGVGVFEGPITASLGTDGSHYKSNWAKAAA 300
DB 241 TIAQANCPYVTVKMSKGAADATAQAKRGVGVFEGPITASLGTDGSHYKSNWAKAAA 300
QY 301 FVTSPPVNPDPPTADHLTCLSSGDLQVTSAGHCTFTTAQKAVGKDNFALIPETNGIEE 360
DB 301 FVTSPPVNPDPPTADHLTCLSSGDLQVTSAGHCTFTTAQKAVGKDNFALIPETNGIEE 360
QY 301 FVTSPPVNPDPPTADHLTCLSSGDLQVTSAGHCTFTTAQKAVGKDNFALIPETNGIEE 360
DB 301 FVTSPPVNPDPPTADHLTCLSSGDLQVTSAGHCTFTTAQKAVGKDNFALIPETNGIEE 360
QY 361 RSMVWEKCVASGKMDNEFVAVTSTNAKIFNYPKRGVAVGSDADLVWPNKATKII 420
DB 361 RSMVWEKCVASGKMDNEFVAVTSTNAKIFNYPKRGVAVGSDADLVWPNKATKII 420
QY 361 RSMVWEKCVASGKMDNEFVAVTSTNAKIFNYPKRGVAVGSDADLVWPNKATKII 420
DB 361 RSMVWEKCVASGKMDNEFVAVTSTNAKIFNYPKRGVAVGSDADLVWPNKATKII 420
QY 421 SAKTHNLNVEYNIFEGVRCGAPAVVISQGRVALEDGKMEVTPGAGRFVPRKTPPDEVYK 480
DB 421 SAKTHNLNVEYNIFEGVRCGAPAVVISQGRVALEDGKMEVTPGAGRFVPRKTPPDEVYK 480
QY 421 SAKTHNLNVEYNIFEGVRCGAPAVVISQGRVALEDGKMEVTPGAGRFVPRKTPPDEVYK 480
DB 421 SAKTHNLNVEYNIFEGVRCGAPAVVISQGRVALEDGKMEVTPGAGRFVPRKTPPDEVYK 480
QY 481 RIKARNRLAEITHGVPRGLYDGPVHEVMVPAKPGSGAPARASCPGKISVPPVRLNHQSGFS 540
DB 481 RIKARNRLAEITHGVPRGLYDGPVHEVMVPAKPGSGAPARASCPGKISVPPVRLNHQSGFS 540
QY 481 RIKARNRLAEITHGVPRGLYDGPVHEVMVPAKPGSGAPARASCPGKISVPPVRLNHQSGFS 540
DB 481 RIKARNRLAEITHGVPRGLYDGPVHEVMVPAKPGSGAPARASCPGKISVPPVRLNHQSGFS 540
QY 541 LSGSQADHIIARRTAQKIMAPPGGRSNTLSLS 572
DB 541 LSGSQADHIIARRTAQKIMAPPGGRSNTLSLS 572
QY 540 LSGTQVDEGV-RSASKRIVAPPGGRSNTLSLS 570
DB 540 LSGTQVDEGV-RSASKRIVAPPGGRSNTLSLS 570

RESULT 11
DPY3_XENLA
ID DPY3_XENLA STANDARD; PRT; 571 AA.
AC OI3022:
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE DIHYDROPYRIMIDINASE RELATED PROTEIN-3 (DRP-3) (NEURAL SPECIFIC PROTEIN
DE 1).
GN NSPL.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;

QY 121 EWRERADSAACDYSLVHVDITRWHSIEKELEALVKEGVNSFLVPMAYKDRCCQSDSQ 180
DB 121 EWRERADSAACDYSLVHVDITRWHSIEKELEALVKEGVNSFLVPMAYKDRCCQSDSQ 180
QY 181 MYEFTSIIRDLGALAAQVHAENGDIIEVEEQKRLLELGTGEGHVLSPHEVEAEAVYAV 240
DB 181 MYEFTSIIRDLGALAAQVHAENGDIIEVEEQKRLLELGTGEGHVLSPHEVEAEAVYAV 240
QY 241 TIAQANCPYVTVKMSKGAADATAQAKRGVGVFEGPITASLGTDGSHYKSNWAKAAA 300
DB 241 TIAQANCPYVTVKMSKGAADATAQAKRGVGVFEGPITASLGTDGSHYKSNWAKAAA 300
QY 301 FVTSPPVNPDPPTADHLTCLSSGDLQVTSAGHCTFTTAQKAVGKDNFALIPETNGIEE 360
DB 301 FVTSPPVNPDPPTADHLTCLSSGDLQVTSAGHCTFTTAQKAVGKDNFALIPETNGIEE 360
QY 361 RSMVWEKCVASGKMDNEFVAVTSTNAKIFNYPKRGVAVGSDADLVWPNKATKII 420
DB 361 RSMVWEKCVASGKMDNEFVAVTSTNAKIFNYPKRGVAVGSDADLVWPNKATKII 420
QY 421 SAKTHNLNVEYNIFEGVRCGAPAVVISQGRVALEDGKMEVTPGAGRFVPRKTPPDEVYK 480
DB 421 SAKTHNLNVEYNIFEGVRCGAPAVVISQGRVALEDGKMEVTPGAGRFVPRKTPPDEVYK 480
QY 481 RIKARNRLAEITHGVPRGLYDGPVHEVMVPAKPGSGAPARASCPGKISVPPVRLNHQSGFS 540
DB 481 RIKARNRLAEITHGVPRGLYDGPVHEVMVPAKPGSGAPARASCPGKISVPPVRLNHQSGFS 540
QY 541 LSGSQADHIIARRTAQKIMAPPGGRSNTLSLS 572
DB 540 LSGTQVDEGV-RSASKRIVAPPGGRSNTLSLS 570

RESULT 10
DPY3_HUMAN
ID DPY3_HUMAN STANDARD; PRT; 570 AA.
AC Q14195; Q93012;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE DIHYDROPYRIMIDINASE RELATED PROTEIN-3 (DRP-3) (UNC-33-LIKE
DE PHOSPHOPROTEIN) (ULIP PROTEIN).
GN DPY3L3 OR ULIP OR DRP3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=97128821; PubMed=8973361;
RA Hamajima N., Matsuda K., Sakata S., Tamaki N., Sasaki M., Nonaka M.;
RT "A novel gene family defined by human dihydropyrimidinase and three
RT related proteins with differential tissue distribution.";
RL Gene 180:157-163(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=9727371; PubMed=9115293;
RA Gaetano C., Matsuo T., Thiele C.J.;
RT "Identification and characterization of a retinoic acid-regulated
RT human homologue of the unc-33-like phosphoprotein gene (hulip) from
RT neuroblastoma cells.";
RL J. Biol. Chem. 272:12195-12201(1997).
CC -1- SUBCELLULAR LOCATION: CYTOSOL.
CC -1- TISSUE SPECIFICITY: MAINLY IN HEART AND SKELETAL MUSCLE. ALSO
CC EXPRESSED STRONGLY IN FETAL BRAIN AND SPINAL CORD.
CC -1- SIMILARITY: BELONGS TO THE DIHYDROPYRIMIDINASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```
CC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Engel E.R., Lepperdinger G., Richter K.;
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE DEHYDROXYRIMIDINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Y13069; CAA73509.1; -
DR InterPro: IPR002195; -
DR Pfam: PF00744; Dihydroorotase; 1.
SQ SEQUENCE 571 AA; 62099 MW; 30A841522E27743D CRC64;

Query Match 72.8%; Score 2172.5; DB 1; Length 571;
Best Local Similarity 69.8%; Pred. No. 1.3e-149;
Matches 399; Conservative 85; Mismatches 87; Indels 1; Gaps 1;

QY 1 MSFGKKSIPRITSDRLIRGGIRVNDQSFYADVHVEDGLIKQIGENLIVPGGIHTIDA 60
DB 1 MSYQKKNIPRITSERLLIKGGIRVNDQSFYADIWEDGLIKQIGENLIVPGGVKTIEA 60
QY 61 HGLMVLPGGVVHTRLOMPVLGMPADDFCQGTAAALAGTTMILDHVPDGTGVSLLAA 120
DB 61 NGKMWIPGGIDVHTRLOMPYRGTMTVDLQGTAAALAGTTMIVDHVPEASLTFEAF 120
QY 121 EOWRERADSAACCDYSLHVDITRWHESTKEEALVKEKGVNSFLVFMAYKDRCCQSDSQ 180
DB 121 EKREWADGKTCCDYSLVHDIHWSVSVQEVETLVKQGVNSFMVYMAKDYQMSNTE 180
QY 181 MYEIFSIRDLGALQAQVHAENGDIVEEQRKLELIGITGPEGHVLSHPPEEAEAVYRAV 240
DB 181 LYEIFTFLGLGALQAQVHAENGDIQAQVHAENGDIQAQVHAENGDIQAQVHAENGDI 240
QY 241 TIAQANCPLVYTKVMSKGAADATQAQKRRGVVVFGEPTITASLGTGTHYWSKNWAKAAA 300
DB 241 ITASQTCNPLYTKVMSKSDVLIISQARKKGVYVFGPEITASLGTGTHYWSKNWAKAAA 300
QY 301 FVTSPPVNDPTADHLTCLLSGDLQVTSAGHCTFTTAQKAVGKDNFALIEGTINGIEE 360
DB 301 FVTSPPSPDPTPDYINSLASGLQVTSAGHCTFTTAQKAVGKDNFALIEGTINGIEE 360
QY 361 RMSWYWEKCVASCKMDENEFVAVTSTNAKIFNTPYRGRVAVGSDADLVINPKATKII 420
DB 361 RMSVLDKSVATGRKMDENQFVSVTSTNAKIFNLPYRGRVAVGSDADLVINPKATKII 420
QY 421 SAKTHNLVAYNIFEGVRCRGAAPVVISQGRVALDQKMEVTPGAGRFVPRKTPDFDYK 480
DB 421 SAKSHSAAEYNIPEGMEELRGAPLVICQKIMMEDGTLHWQGRFIPCSPPDYVYK 480
QY 481 RIKARNRLAETHVPRGLYDGPVHEVWVPAKPGSGAPARASCPGKISVPPVNLHQSGFS 540
DB 481 RIKARTKMAELHVPGRMGYDGPVLDASTPRAGTPAGTKGSPK-QTPPVNRLHHSAFS 539
QY 541 LSGQADPHIARRTAQKIMAPPGGRSNTITSL 572
DB 540 LAGNOGDESGVRSARRIVAPLGAPNITSL 571

RESULT 12
DPY1_MOUSE
ID DPY1_MOUSE STANDARD; PRT; 572 AA.
AC P97427; O08554; O35097;
DT 15-JUL-1998 (Rel. 36, Created)
```

```
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE DIHYDROXYRIMIDINASE RELATED PROTEIN-1 (DRP-1) (COLLAPSPIN RESPONSE
DE MEDIATOR PROTEIN 1) (CRMP-1) (ULIP3 PROTEIN).
GN CRMP1 OR DPYSLI OR ULIP3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97262103; PubMed=9107681;
RA Cohen-Salmon M., Crozet F., Rebillard G., Petit C.;
RT "Cloning and characterization of the mouse collapsin response
RT mediator protein-1, Crmp1";
RL Mamm. Genome 8:349-351(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ICR; TISSUE=Brain;
RX MEDLINE=98314496; PubMed=9652388;
RA Byk T., Ozon S., Sobel A.;
RT "The Ulip family phosphoproteins -- common and specific properties.";
RL Eur. J. Biochem. 254:14-24(1998).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RA Hamajima N., Kato Y., Kowaki M., Wada Y., Sasaki M., Nonaka M.;
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE DEHYDROXYRIMIDINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U72875; AAB39703.1; -
DR EMBL: Y09080; CAA70300.1; -
DR EMBL: AB006714; BAA21887.1; -
DR MGD; MGI:107793; Crmp1.
DR InterPro: IPR002195; -
DR Pfam: PF00744; Dihydroorotase; 1.
FT CONFLICT 338 338 T -> I (IN REF. 2).
FT CONFLICT 476 476 E -> K (IN REF. 3).
FT CONFLICT 489 489 F -> S (IN REF. 3).
FT CONFLICT 520 520 K -> E (IN REF. 2).
SQ SEQUENCE 572 AA; 62167 MW; FE17DDCD735CAF8F CRC64;

Query Match 72.3%; Score 2157; DB 1; Length 572;
Best Local Similarity 69.2%; Pred. No. 1.7e-148;
Matches 395; Conservative 85; Mismatches 91; Indels 0; Gaps 0;

QY 1 MSFGKKSIPRITSDRLIRGGIRVNDQSFYADVHVEDGLIKQIGENLIVPGGIHTIDA 60
DB 1 MSYQKKNIPRITSERLLIKGGIRVNDQSFYADIWEDGLIKQIGENLIVPGGVKTIEA 60
QY 61 HGLMVLPGGVVHTRLOMPVLGMPADDFCQGTAAALAGTTMILDHVPDGTGVSLLAA 120
DB 61 NGKMWIPGGIDVHTRLOMPYRGTMTVDLQGTAAALAGTTMIVDHVPEASLTFEAF 120
QY 121 EOWRERADSAACCDYSLHVDITRWHESTKEEALVKEKGVNSFLVFMAYKDRCCQSDSQ 180
DB 121 EKREWADGKTCCDYSLVHDIHWSVSVQEVETLVKQGVNSFMVYMAKDYQMSDQ 180
QY 181 MYEIFSIRDLGALQAQVHAENGDIVEEQRKLELIGITGPEGHVLSHPPEEAEAVYRAV 240
DB 181 LYEIFTFLGLGALQAQVHAENGDIQAQVHAENGDIQAQVHAENGDIQAQVHAENGDI 240
QY 241 TIAQANCPLVYTKVMSKGAADATQAQKRRGVVVFGEPTITASLGTGTHYWSKNWAKAAA 300
```

Tue Jul 31 13:08:20 2001

0:

Matches 393; Conservative 87; Mismatches 91; Indels 0; Gaps 0;

```

1 MSFOGKKSIPRITSDRLIRGRIVNDQSFYADVHVDEGLIKOIGENLIVPGIHTIDA 60
1 MSHOGKKSIPHTSDRLIRGRIVNDQSFYADVHVDEGLIKOIGENLIVPGVKTIEA 60
61 HGLVLPVGVVHYFLOMPVMTADDPCOGTKAALAGGTTMILDHVFPDPTGVSLAAY 120
61 NGRMVPVIGSDIVNTYLPQPSQMTSADDFGOTRAALAGGTTMIDHVPVPGSSLTSE 120
121 EOWHERASAAACCDYSLHVDITRWHESESKEEALVKEGVNSFLVMAYKDRCCSDSQ 180
121 EKHIEAADTKCCDYSLHVDITRWHESESKEEALVKEGVNSFLVMAYKDRCCSDSQ 180
181 MYEFSIIRDLGALAOVHAENGDIVEEQKRLLELGTGPEGHVLSHPPEVEAEAVYRAV 240
181 LYEAFTFLKGLGAVILVHAENGDLIAOEKRLLELGTGPEGHVLSHPPEVEAEAVYRAV 240
241 TIAKQANCPVYTKVMSKGAADIAOAKRRGVVVFGEPTASLGTGSHYMSKNWAKAAA 300
241 AIAGRINCPVYTKVMSKGAADIAOAKRRGVVVFGEPTASLGTGSHYMSKNWAKAAA 300
301 FVTSPPVNPDPPTADHLTCLSSGDLQVTSAGHCTTTAKAVGKDNFALIEPENGIEE 360
301 FVTSPLSPDPPTPDYLTSLACGDLQVTSAGHCTTTAKAVGKDNFALIEPENGIEE 360
361 RMSVMWEKCVASGKMDENEFVAVTSTNAAKIFNYPKRGVAVGSDADLVINPRATKII 420
361 RMTVVWDKAVATGKMDENEFVAVTSTNAAKIFNYPKRGVAVGSDADLVINPRATKII 420
421 SAKTHNLNVEYNIFEGVEGRCGAPAVVISQGRVALEDGKMFVTPGAGRFVPRKTFPDPVYK 480
421 TAKSHKSTVEYNIFEGMECHGSPVVISQGRVALEDGKMFVTPGAGRFVPRKTFPDPVYK 480
481 RIKARNRLAEIHGVRGLYDGPVHEVMVPAKPGSAPARASCPGKISVPPVRNLHOSGFS 540
481 RVRIRSKVFLHSVSRGMYDGPVVEVPATPKHAAPASAKSSPSKHPPIRNLHOSNFS 540
541 LSGSQADDDHIAARTAKIMAPGGRSNITSL 571
541 LSGAQIDDDNPRRTGHRIVAPPGGRSNITSL 571

```

RESULT 14
ID DPYL_HUMAN STANDARD; PRT; 572 AA.
AC Q14194; Q13024;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE DIHYDROPYRIMIDINASE RELATED PROTEIN-1 (DRP-1) (COLLAPSPIN RESPONSE
DE MEDIATOR PROTEIN 1) (CRMP-1).
GN CRMP1 CR DPYSL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=97128821; PubMed=8973361;
RA Hamajima N., Matsuda K., Sakata S., Tamaki N., Sasaki M., Nonaka M.;
RT "A novel gene family defined by human dihydropyrimidinase and three
RL related proteins with differential tissue distribution."
RN [2]
RP SEQUENCE OF 64-572 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=95364923; PubMed=7637782;
RA Goshima Y., Nakamura F., Strittmatter P., Strittmatter S.M.;
RT "Collapsin-induced growth cone collapse mediated by an intracellular
RL protein related to UNC-33."
RN Nature 376:509-514 (1995).

```

241 AIAGRINCPVYTKVMSKGAADIAOAKRRGVVVFGEPTASLGTGSHYMSKNWAKAAA 300
301 FVTSPPVNPDPPTADHLTCLSSGDLQVTSAGHCTTTAKAVGKDNFALIEPENGIEE 360
301 FVTSPLSPDPPTPDYLTSLACGDLQVTSAGHCTTTAKAVGKDNFALIEPENGIEE 360
361 RMSVMWEKCVASGKMDENEFVAVTSTNAAKIFNYPKRGVAVGSDADLVINPRATKII 420
361 RMTVVWDKAVATGKMDENEFVAVTSTNAAKIFNYPKRGVAVGSDADLVINPRATKII 420
421 SAKTHNLNVEYNIFEGVEGRCGAPAVVISQGRVALEDGKMFVTPGAGRFVPRKTFPDPVYK 480
421 TAKSHKSTVEYNIFEGMECHGSPVVISQGRVALEDGKMFVTPGAGRFVPRKTFPDPVYK 480
481 RIKARNRLAEIHGVRGLYDGPVHEVMVPAKPGSAPARASCPGKISVPPVRNLHOSGFS 540
481 RVRIRSKVFLHSVSRGMYDGPVVEVPATPKHAAPASAKSSPSKHPPIRNLHOSNFS 540
541 LSGSQADDDHIAARTAKIMAPGGRSNITSL 571
541 LSGAQIDDDNPRRTGHRIVAPPGGRSNITSL 571

```

RESULT 13
ID DPYL_RAT STANDARD; PRT; 572 AA.
AC Q62950; P70546;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE DIHYDROPYRIMIDINASE RELATED PROTEIN-1 (DRP-1) (COLLAPSPIN RESPONSE
DE MEDIATOR PROTEIN 1) (CRMP-1).
GN CRMP1 OR DPYSL.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=96424532; PubMed=8815901;
RA Wang L., Strittmatter S.M.;
RT "A family of rat CRMP genes is differentially expressed in the
RL nervous system."
RN J. Neurosci. 16:6197-6207 (1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=Brain;
RA Quach T.T., Honnorat J., Aguerre M., Belin M.F., Kolattukudy P.E.,
RN Antoine J.C.;
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
CC 1- SUBCELLULAR LOCATION: CYTOPLASMIC (POTENTIAL).
CC 1- SIMILARITY: BELONGS TO THE DIHYDROPYRIMIDINASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; U52102; AAB03280.1; -;
CC EMBL; U52095; AAB07042.1; -;
CC InterPro; IPR002195; -;
CC Pfam; PF00744; Dihydroorotase; 1.
CC CONFLICT 3 H -> Y (IN REF. 2).
CC FT SEQUENCE 572 AA; 62195 MW; ED63BD8C751CCPDF CRC64;
CC
Query Match 72.1%; Score 2152; DB 1; Length 572;
Best Local Similarity 68.8%; Pred. No. 4e-148;

```

CC CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (POTENTIAL).
CC CC -1- TISSUE SPECIFICITY: BRAIN.
CC CC -1- SIMILARITY: BELONGS TO THE DEHYDROPYRIMIDINASE FAMILY.
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC CC at the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
CC CC EMBL; D78012; BAA11190.1;
CC CC EMBL; U17278; AAA93201.1;
CC CC MIM; 602462;
CC CC InterPro; IPR002195;
CC CC Pfam; PF00744; Dihydroorotase; 1.
CC CC CONFLICT 504 504 Y -> H (IN REF. 2).
CC CC SEQUENCE 572 AA; 62183 MW; A5385FCC79328A30 CRC64;
CC CC -----
CC CC Query Match 72.1%; Score 2151; DB 1; Length 572;
CC CC Best Local Similarity 69.0%; Pred. No. 4.7e-148;
CC CC Matches 394; Conservative 85; Mismatches 92; Indels 0; Gaps 0;
CC CC -----
QY 1 MSFGCKSIPRITSDRLIRGRVINDQSFYADVHVEDGLIKOIGENLIVPGGIHTIDA 60
DB 1 MSYQKKSIPHTSDRLIKGRVINDQSLYADVLEDGLIKOIGENLIVPGGVKTI EA 60
QY 61 HCLMVLPGVDVHTRLQMPVLGMPADFCQGTAAALAGGTTMLDHFVPTDGVSLAA 120
DB 1 NGRMVFPGIDVNTYLOKPSQGMTAADFFQGTAAALVGGTTMLDHFVPEGSSLLT 120
QY 121 BQWRERADSAACCDYSLHVDITRWHESIKELEALYKEGVNSFLVPMAYKDCQCSDSQ 180
DB 121 EKHEADATKSCDDYSLHVDITSWIDVREELVLDKGVNSFQVMYKDVYQMSDSQ 180
QY 181 MYEFTSIIRDLGALQVHAENGDIVEEQKRLLELGTITGPEGHVLSHPEVEAEYRAY 240
DB 181 LYEAFTFLKGLGAVILVHAENGDLIAEQKRIEMGITGPEGHVLSHPEVEAEYRAY 240
QY 241 TIAQANCPVYTVKMSKGAADIAOAKRGVVVFCEPTASLGTDGSHYKSNWAKAAA 300
DB 241 TIAGRINCPVYTVKMSKGAADIAOAKRGVVVFCEPTASLGTDGSHYKSNWAKAAA 300
QY 301 FVTSPPVNDPTADHLTCLLSGDLQVTSARCTFTTAQKAVGKDNFALIEGTNGIEE 360
DB 301 FVTSPPVNDPTADHLTCLLSGDLQVTSARCTFTTAQKAVGKDNFALIEGTNGIEE 360
QY 361 RMSWVWEKCVASKMDENEFVAVTSTNAKIFNFKRGVAVGSDADLVINWPKATKII 420
DB 361 RMTVVWDKAVATGKNDENEFVAVTSTNAKIFNFKRGVAVGSDADLVINWPKATKII 420
QY 421 SAKTHNLVYNIIFGVFCRGAPAVVISQGRVLEDDGKMFVTPGAGRVPRKTPDFVYK 480
DB 421 TAKSHKSAVEYNIFEGMECHSPLVVISQGRVLEDDGKMFVTPGAGRVPRKTPDFVYK 480
QY 481 RIKARNRLAEIHGVPRLGLYDGPVHEVYPAKPGSGAPARASCPOKISVPPVNLHQSGFS 540
DB 481 RYKIRNKVFGLOGVSRGMDGVPVEVPATPKYATPAPSAKSSPSKHPQPPVNLHQSGFS 540
QY 541 LSGAQDIDNNRRTRGHRIVAPPGRSNTSL 571
DB 541 LSGAQDIDNNRRTRGHRIVAPPGRSNTSL 571
CC CC -----
CC CC RESULT 15
CC CC ID DPVS_HUMAN
CC CC AC Q14117;
CC CC DT 15-JUL-1998 (Rel. 36, Created)
CC CC DT 15-JUL-1998 (Rel. 36, Last sequence update)
CC CC DT 15-DEC-1998 (Rel. 37, Last annotation update)

```

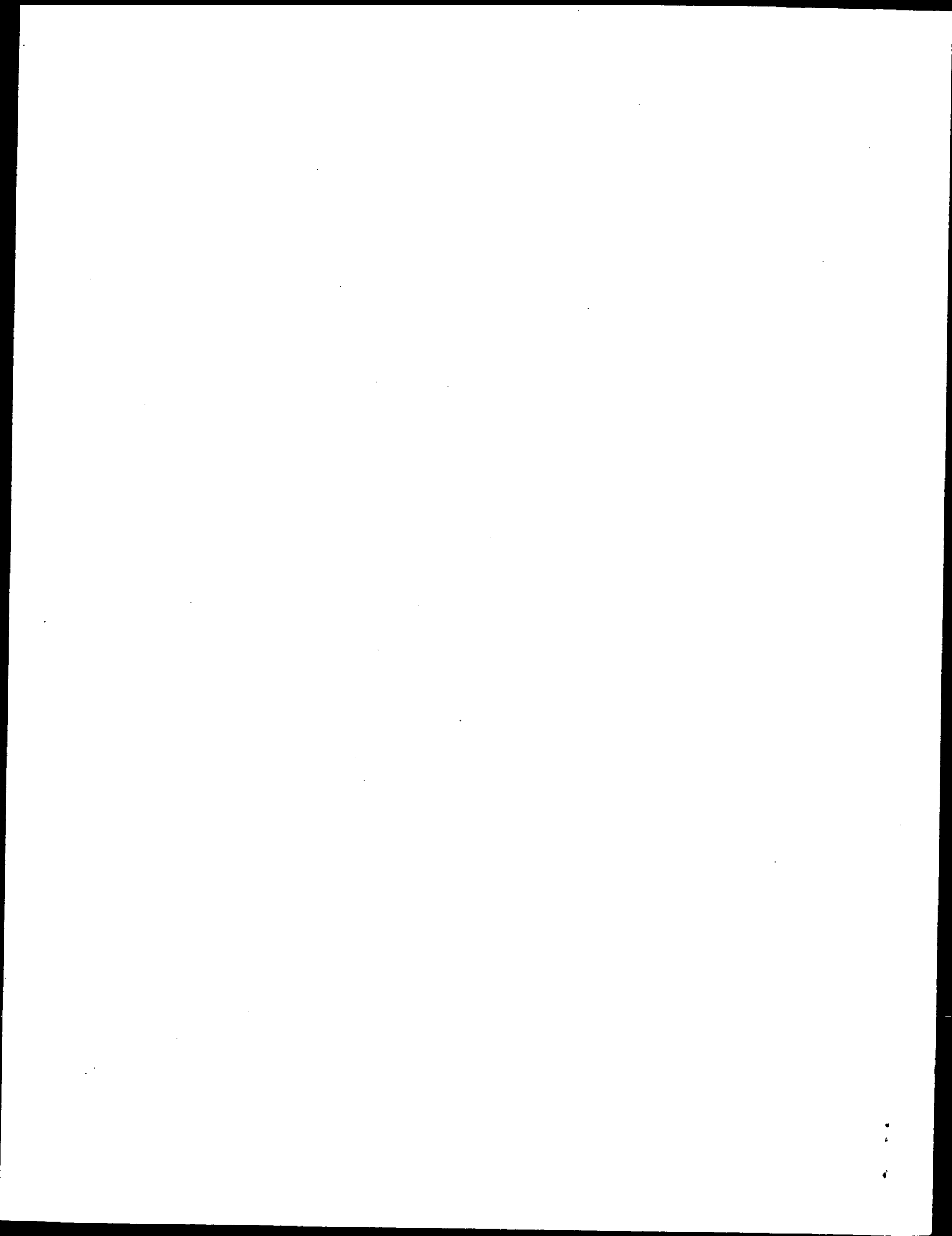
```

DE GN DIHYDROPYRIMIDINASE (EC 3.5.2.2) (DHPASE) (HYDANTOINASE) (DHP).
OS DPVS.
OS Homo sapiens (Human).
CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
CC NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=97128821; PubMed=8973361;
RA Hamajima N., Matsuda K., Sakata S., Tamaki N., Sasaki M., Nonaka M.;
RT "A novel gene family defined by human dihydropyrimidinase and three
RL related proteins with differential tissue distribution.";
RN [2]
RP SEQUENCE FROM N.A. AND VARIANTS DHP DEFICIENCY.
RX MEDLINE=98386066; PubMed=9718352;
RA Hamajima N., Kouchi K., Vreken P., Matsuda K., Sumi S., Imaeda M.,
RA Ohba S., Kidouchi K., Nonaka M., Sasaki M., Tamaki N., Endo Y.,
RA de Abreu R., Rottevell J., van Kullenburg A., van Gennip A.,
RA Togari H., Wada Y.;
RT "Dihydropyrimidinase deficiency: structural organization, chromosomal
RT localization, and mutation analysis of the human dihydropyrimidinase
RT gene.";
RL Am. J. Hum. Genet. 63:717-726(1998).
CC -1- CATALYTIC ACTIVITY: 5,6-DIHYDROURACIL + H(2)O = 3-
CC UREIDOPROPIONATE.
CC -1- TISSUE SPECIFICITY: LIVER AND KIDNEY.
CC -1- DISEASE: DEFECTS IN DPVS ARE THE CAUSE OF AN AUTOSOMAL RECESSIVE
CC DISORDER CHARACTERIZED BY DIHYDROPYRIMIDINURIA AND ASSOCIATED WITH
CC A VARIABLE CLINICAL PHENOTYPE (EPILEPTIC OR CONVULSIVE ATTACKS,
CC DYSMORPHIC FEATURES AND SEVERE DEVELOPMENTAL DELAY, AND CONGENITAL
CC MICROVILLOUS ATROPHY).
CC -1- SIMILARITY: BELONGS TO THE DEHYDROPYRIMIDINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D78011; BAA11189.1;
CC EMBL; AB004678; BAA33067.1;
CC EMBL; AB004669; BAA33067.1; JOINED.
CC EMBL; AB004670; BAA33067.1; JOINED.
CC EMBL; AB004671; BAA33067.1; JOINED.
CC EMBL; AB004672; BAA33067.1; JOINED.
CC EMBL; AB004673; BAA33067.1; JOINED.
CC EMBL; AB004674; BAA33067.1; JOINED.
CC EMBL; AB004675; BAA33067.1; JOINED.
CC EMBL; AB004676; BAA33067.1; JOINED.
CC EMBL; AB004677; BAA33067.1; JOINED.
CC MIM; 222748;
CC InterPro; IPR002195;
CC Pfam; PF00744; Dihydroorotase; 1.
CC KW Hydrolyase; Disease mutation.
CC VARIANT 68 68
CC T -> R (IN DHP DEFICIENCY).
CC Q -> R (IN DHP DEFICIENCY).
CC W -> R (IN DHP DEFICIENCY).
CC G -> R (IN DHP DEFICIENCY).
CC R -> T (IN DHP DEFICIENCY).
CC /FTid=VAR_002267.
CC /FTid=VAR_002268.
CC /FTid=VAR_002269.
CC /FTid=VAR_002270.
CC /FTid=VAR_002271.
CC SEQUENCE 519 AA; 56629 MW; 882E33D7C49D6ECC CRC64;
CC -----
CC Query Match 52.1%; Score 1554; DB 1; Length 519;
CC Best Local Similarity 59.3%; Pred. No. 5.8e-105;

```

Matches	292;	Conservative	69;	Mismatches	127;	Indels	4;	Gaps	1;
Qy	16	RLIRGRIVNDQSFYADYVHVEDGLIKOIGENLIYVG----	GIHTIDAHGLMLVPGGVD	71					
Db	6	RLIRGRVYVNDSEVADYVHVEDGLIKOIGENLIYVG----	GIHTIDAHGLMLVPGGVD	65					
Qy	72	VHTRLOMPVLGMPADDFCOGTRKALAGGTTMLDHFVDPDTGVSLAAYEQWREERADSA	131						
Db	66	THTHMQPFPMGSRSDIDFHQTRKALSGGTTMLDHFVDPDTGVSLAAYEQWREERADSA	125						
Qy	132	CCDYSLHVDITRWHESTKELEALVKEKGVNSFLVPMAYKDRCCSDSQMVEYLFIRDL	191						
Db	126	CCDYSLHVAVTWMSDQYKEEMKILVDQKGVNSFKMFEMAYKDLVYVTDLELYEAFSRCKE	185						
Qy	192	GALAQVHAENGDIIVEEQKRLLELIGITGPEGHVLSHPEEVEAEAVYRAVIAKQANCPLY	251						
Db	186	GATVHAENGDLIAEGAKKMLALGITGPEGHELCPPEAEAEATLRAITIASAVNCPY	245						
Qy	252	VTKYNKGAADAIQAQRGRVYVGEPIITASLGTDGSHYWSKNWAKAAAFVSPVNPDP	311						
Db	246	IVHNSKSAKVIADARRDGKVVYGEPIITASLGTDGTHYWNKEWHHAHHVMGPPLRDP	305						
Qy	312	TTADHLICLLSSGDLQVTSAGHCTFTTAQKAVGKDNFALIPETGNGIEERMSVWKEKVA	371						
Db	306	STPDFLMLNLANDLTTGTDCNTFCQKALGKDDFTKIPNGVNGVEDRMSVWKEKVA	365						
Qy	372	SGKMDENEFVAVTSTNAKIFENFYPRKGRVAVGSDADLVINPKATKIIISAKTHNLNVEY	431						
Db	366	SGKMDENEFVAVTSTNAKIFENFYPRKGRVAVGSDADLVINPKATKIIISAKTHNLNVEY	425						
Qy	432	NIFEGVECRGAPAVVISGRVALEDGKMFVTPGAGRFPVPRKTFPDPFVYKRIKARNRLAEI	491						
Db	426	NIFEGMCHGVPLVTISRKVVYAGVSVTAGDGKFIIPKPFPAEYIYKRIKQDRCTCTP	485						
Qy	492	HGVPRGLYDGPV	503						
Db	486	TPVERAPYKGEV	497						

Search completed: July 30, 2001, 11:42:21
Job time: 140 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2001, 11:37:26 ; Search time 16.35 Seconds
(without alignments)
2664.945 Million cell updates/sec

Title: us-09-367-496-8
Perfect score: 2985
Sequence: 1 MSFQKKIPRITSDRLIR.....RTAQKIMAPPGGRSNTISLS 572

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_68: *
1: pir1: *
2: pir2: *
3: pir3: *
4: pir4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2362	79.1	572	2 S58889	collapsin response
2	2343	78.5	572	2 S49985	dihydropyrimidinas
3	2339	78.4	572	2 JC5317	dihydropyrimidinas
4	2187	73.3	570	2 S55525	ulip protein - mou
5	2186	73.2	570	2 JC5318	dihydropyrimidinas
6	2151	72.1	572	2 JC5316	dihydropyrimidinas
7	1554	52.1	519	2 JC5315	dihydropyrimidinas
8	1548	51.9	519	2 S70581	hypothetical prote
9	1245	41.7	520	2 T20007	hypothetical prote
10	1122	37.6	544	2 T23968	dihydropyrimidinas
11	1074.5	36.0	479	2 H83590	unc-33 protein - C
12	837	28.0	471	1 JC2310	hypothetical prote
13	781.5	26.2	854	2 S33558	hypothetical prote
14	692.5	23.2	467	2 T28685	hypothetical prote
15	672	22.5	465	2 A65071	probable allantoin
16	672	22.5	465	2 F85942	allantoinase BH230
17	325.5	10.9	448	2 E75429	probable allantoin
18	317	10.6	438	2 E83938	probable dihydroor
19	311.5	10.4	453	1 G64782	allantoinase homol
20	296.5	9.9	430	2 B70959	dihydroorotase [im
21	295.5	9.9	446	1 C70016	dihydroorotase (EC
22	295.5	9.9	454	2 A69017	dihydroorotase pyr
23	259.5	8.7	422	2 T45151	hypothetical prote
24	242	8.1	430	2 T45151	pyrimidine synth
25	241.5	8.1	422	2 C70370	hypothetical prote
26	233.5	7.8	428	2 B83967	hypothetical prote
27	231.5	7.8	2198	2 T20371	hypothetical prote
28	229.5	7.7	2242	2 A57541	hypothetical prote
29	229	7.7	296	2 H85550	hypothetical prote

ALIGNMENTS

RESULT 1

S58889

collapsin response mediator protein, 62K - chicken

C:Species: Gallus gallus (chicken)

C:Date: 15-Feb-1996 #sequence_revision 01-Mar-1996 #text_change 02-Sep-2000

C:Accession: S58889

R:Goshima, Y.; Nakamura, F.; Strittmatter, P.; Strittmatter, S.M.

A:Title: Collapsin-induced growth cone collapse mediated by an intracellular protein

A:Reference number: S58889; MUID:95364923

A:Accession: S58889

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-572 <GOS>

A:Cross-references: EMBL:U17277; NID:9882146; PIDN:AAA93200.1; PID:9882147

C:Superfamily: allantoinase; Bacillus dihydroorotase homology

Query Match 79.1%; Score 2362; DB 2; Length 572;
Best Loca.: Similarity 76.4%; Pred. No. 8e-166;
Matches 436; Conservative 68; Mismatches 67; Indels 0; Gaps 0;

QY	1	MSFQKKIPRITSDRLIRGRVINDQSFADVHVEDGLIKGICENLIVPGGIHTDA	60
DB	1	MSFQKKIPRITSDRLIRGRVINDQSFADVHVEDGLIKGICENLIVPGGIHTDA	60
QY	61	HGLMVLPGVDVHTRLOMPVLGTPADDFCQGTKAALAGGTTMILDRVFPDTCVSLAAY	120
DB	61	HGRMVLPGVDVHTRLOMPVLGTPADDFCQGTKAALAGGTTMILDRVFPDTCVSLAAY	120
QY	121	EWRERADSAACDYSILHVDITRWHSIKLEALYKEGVNSFLVFMAYKRCQSDSQ	180
DB	121	DQREWADSKSCDYSILHVDITRWHSIKLEALYKEGVNSFLVFMAYKRCQSDSQ	180
QY	181	MYEFTSIIRDLGALAOVHAENGDIIEEQKRLLELGTGPEGHVLSHPHEVEAEVYRV	240
DB	181	IYEVLSVIRDIGTAQVHAENGDIIEEQKRLLELGTGPEGHVLSHPHEVEAEVYRV	240
QY	241	TIQAQNCPLYVTYKMSKGAADAIAQARRGVVVFGEPIITASLGTDGSHVSKNNAKAAA	300
DB	241	TIQAQNCPLYVTYKMSKGAADAIAQARRGVVVFGEPIITASLGTDGSHVSKNNAKAAA	300
QY	301	FVTSPPVNPDPPTADHITLCLLSSGDLQVTSAGCTFTTAQKAVGKDNFALIPCTNGIEE	360
DB	301	FVTSPPVNPDPPTADHITLCLLSSGDLQVTSAGCTFTTAQKAVGKDNFALIPCTNGIEE	360
QY	361	PMVMWEKCVASGKMDNEFVAVTSTNAKIFNFPYPRKRVAVGSDADLVINPDKATKII	420
DB	361	PMVMWEKCVASGKMDNEFVAVTSTNAKIFNFPYPRKRVAVGSDADLVINPDKATKII	420
QY	421	SAKTHNLNVEYNIFEGVECRGAPAVVISQGRVALEDGKMFVTPGAGRFVPRKTPDFVYK	480
DB	421	SAKTHNLNVEYNIFEGVECRGAPAVVISQGRVALEDGKMFVTPGAGRFVPRKTPDFVYK	480

Db 421 SAKTHNISLEYNIFEGMECRGSPVVISQKIVLELDGNLHVTEGSGRYIPKRPFPDFVYK 480
 QY 481 RIKARNRLAEIHGVRGLYDGPVHEVWVPKPGSGAPARASCPGKISVPPVNLHQSFGS 540
 Db 481 RIKARSRLAELRGVRLGGLYDGPVCEVSVTPKVTVPASSAKTSPAKQAPPPVNLHQSFGS 540
 QY 541 LSGSQADDDHARTTAQKIMAPPGGRSNTSL 571
 Db 541 LSGAQIDDDNIPRRTTQRIVAPPGGRANITSL 571

RESULT 2
 S49985
 dihydropyrimidinase-related protein 2 [similarity] - rat
 N:Alternate names: collapsin response mediator protein 2; TOAD-64
 C:Species: Rattus norvegicus (Norway rat)
 C:Date: 13-Jan-1995 #sequence_revision 12-May-1995 #text_change 02-Sep-2000
 C:Accession: A59280; S49985
 R:Minturn, J.E.; Fryer, H.J.; Geschwind, D.H.; Hockfield, S.
 J. Neurosci. 15, 6757-6766, 1995
 A:Title: TOAD-64, a gene expressed early in neuronal differentiation in the rat, is related to the human dihydropyrimidinase-related protein 2 (DRP-2)
 A:Reference number: A59280; MUID:96033765
 A:Accession: A59280
 A:Status: Preliminary
 A:Molecule type: mRNA
 A:Residues: 1-572 <MIN>
 A:Cross-references: GB:246882; NID:G599965; PIDN:CAA86981.1; PID:G599966
 A:Experimental source: Strain Sprague-Dawley; clone TOAD64 (turned on after division, 64 hours after birth)
 A:Note: Submitted to the EMBL Data Library, November 1994
 A:Note: In Genbank entry R1TOAD64, release 113.0, the source is designated as Rattus rattus
 C:Superfamily: allantoicase; Bacillus dihydroorotase homology

Query Match 78.5%; Score 2343; DB 2; Length 572;
 Best Local Similarity 75.8%; Pred. No. 2e-164;
 Matches 433; Conservative 70; Mismatches 68; Indels 0; Gaps 0;

QY 1 MSFGKKSIPRITSDDLIRGGIRVNDQSFYADVHVEDGLIKOIGENLIVPGGIHTIDA 60
 Db 1 MSYQGNIPRITSDDLIRGGIRVNDQSFYADVHVEDGLIKOIGENLIVPGGIHTIDA 60
 QY 61 HGLWLPGGVDVHTRIQMPVLGMPADDFCOGTAALAGGTTMIIDHVPDPTGVSLAAY 120
 Db 61 HSRWIPGGIDVHTRIQMPDQGMTSADDFCOGTAALAGGTTMIIDHVPDPTGVSLAAY 120
 QY 121 EOWREADSAACDYSLVHVDITRWHSIKEELEALVKRGVNSFLVFMAYKDRCCSDSQ 180
 Db 121 DQREWADSKCCDYSLVHVDITRWHSIKEELEALVKRGVNSFLVFMAYKDRCCSDSQ 180
 QY 181 MYEIFSIIIDLGALAQVHAENGDIIVEEOKRLLELIGITGPEGHVLSHPPEVEAEAVYRAV 240
 Db 181 IYEVLSVIRIDGALAQVHAENGDIIVEEOKRLLELIGITGPEGHVLSHPPEVEAEAVYRAV 240
 QY 241 TIAQANCPPLYTKVMSKGAADAIAQAKRRGVVVFCEPITASLTGDSHYHYSKNWAKAAA 300
 Db 241 TIANQTNCPPLYTKVMSKGAADAIAQAKRRGVVVFCEPITASLTGDSHYHYSKNWAKAAA 300
 QY 301 FVTSPPVNDPTDHLTCLSSGLDQVTSAGHCTFTTAQAVGKDNFALLIPEGTNGTEE 360
 Db 301 FVTSPPVNDPTDHLTCLSSGLDQVTSAGHCTFTTAQAVGKDNFALLIPEGTNGTEE 360
 QY 361 RMSWVWEKCVASGKMDENEFVAVTSTNAKIFNYPKRGVAVGSDADLVINPDKATKII 420
 Db 361 RMSVWVWAVTGVKMDENEFVAVTSTNAKIFNYPKRGVAVGSDADLVINPDKATKII 420
 QY 421 SAKTHNLNVEYNIFEGVECRGAPAVVISQGRVALEDGKMFVTPGAGREVPRKFTPPDVYK 480
 Db 421 SAKTHNSALEYNIFEGVECRGAPAVVISQGRVALEDGKMFVTPGAGREVPRKFTPPDVYK 480
 QY 481 RIKARNRLAEIHGVRGLYDGPVHEVWVPKPGSGAPARASCPGKISVPPVNLHQSFGS 540
 Db 481 RIKARSRLAELRGVRLGGLYDGPVCEVSVTPKVTVPASSAKTSPAKQAPPPVNLHQSFGS 540

QY 541 LSGSQADDDHARTTAQKIMAPPGGRSNTSL 571
 Db 541 LSGAQIDDDNIPRRTTQRIVAPPGGRANITSL 571

RESULT 3
 JC5317
 dihydropyrimidinase-related protein 2 - human
 N:Alternate names: collapsin response mediator protein 2
 C:Species: Homo sapiens (man)
 C:Date: 01-May-1997 #sequence_revision 23-Aug-1997 #text_change 02-Sep-2000
 C:Accession: JC5317; S58891; JC7150
 R:Hamaajima, N.; Matsuda, K.; Sakata, S.; Tamaki, N.; Sasaki, M.; Nonaka, M.
 Gene 180, 157-163, 1996
 A:Title: A novel gene family defined by human dihydropyrimidinase and three related proteins
 A:Reference number: JC5315; MUID:97128821
 A:Accession: JC5317
 A:Status: nucleic acid sequence not shown
 A:Molecule type: mRNA
 A:Residues: 1-572 <HAM>
 A:Cross-references: DDBJ:D78013; NID:G1330239; PIDN:BAAL1191.1; PID:G1330240
 A:Experimental source: fetal brain
 R:Goshima, Y.; Nakamura, F.; Strittmatter, P.; Strittmatter, S.M.
 Nature 376, 509-514, 1995
 A:Title: Collapsin-induced growth cone collapse mediated by an intracellular protein
 A:Reference number: S58889; MUID:95364923
 A:Accession: S58891
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: mRNA
 A:Residues: 1-572 <GOS>
 A:Cross-references: EMBL:U17279; NID:G1244399; PIDN:AAA93202.1; PID:G1244400
 R:Kitamura, K.; Takayama, M.; Hamaajima, N.; Nakanishi, M.; Sasaki, M.; Endo, Y.; Take
 DNA Res. 6, 291-297, 1999
 A:Title: Characterization of the human dihydropyrimidinase-related protein 2 (DRP-2)
 A:Reference number: JC7150; MUID:20039612
 A:Accession: JC7150
 A:Status: preliminary
 A:Molecule type: nucleic acid
 A:Residues: 1-572 <KIT>
 A:Cross-references: DDBJ:AB020764; PIDN:BA88422.1
 C:Comment: This protein is the homologue of chicken 62K collapsin response mediator p
 C:Genetics:
 A:Gene: GDB:DPYSL2; DHPRP2; DRP-2; CRMP2
 A:Cross-references: GDB:5496401
 A:Map position: 8p22-8p21
 C:Superfamily: allantoicase; Bacillus dihydroorotase homology

Query Match 78.4%; Score 2339; DB 2; Length 572;
 Best Local Similarity 75.5%; Pred. No. 4e-164;
 Matches 431; Conservative 72; Mismatches 68; Indels 0; Gaps 0;

QY 1 MSFGKKSIPRITSDDLIRGGIRVNDQSFYADVHVEDGLIKOIGENLIVPGGIHTIDA 60
 Db 1 MSYQGNIPRITSDDLIRGGIRVNDQSFYADVHVEDGLIKOIGENLIVPGGIHTIDA 60
 QY 61 HGLWLPGGVDVHTRIQMPVLGMPADDFCOGTAALAGGTTMIIDHVPDPTGVSLAAY 120
 Db 61 HSRWIPGGIDVHTRIQMPDQGMTSADDFCOGTAALAGGTTMIIDHVPDPTGVSLAAY 120
 QY 121 EOWREADSAACDYSLVHVDITRWHSIKEELEALVKRGVNSFLVFMAYKDRCCSDSQ 180
 Db 121 DQREWADSKCCDYSLVHVDITRWHSIKEELEALVKRGVNSFLVFMAYKDRCCSDSQ 180
 QY 181 MYEIFSIIIDLGALAQVHAENGDIIVEEOKRLLELIGITGPEGHVLSHPPEVEAEAVYRAV 240
 Db 181 IYEVLSVIRIDGALAQVHAENGDIIVEEOKRLLELIGITGPEGHVLSHPPEVEAEAVYRAV 240
 QY 241 TIAQANCPPLYTKVMSKGAADAIAQAKRRGVVVFCEPITASLTGDSHYHYSKNWAKAAA 300
 Db 241 TIANQTNCPPLYTKVMSKGAADAIAQAKRRGVVVFCEPITASLTGDSHYHYSKNWAKAAA 300

QY 421 SAKTHNLNVEYNIFEGVPCRCAPAVVISQGRVALEDGKMFVTPGAGRFVPRKTPFPDFVYK 480
 Db 421 SAKTHNLNVEYNIFEGVPCRCAPAVVISQGRVALEDGKMFVTPGAGRFVPRKTPFPDFVYK 480
 QY 481 RIKARNRLAEIHGVPRGLYDGPVHEVMPAKPGCAPARASCPGKISVPPVRNLHQSGFS 540
 Db 481 RIKARNRLAEIHGVPRGLYDGPVHEVMPAKPGCAPARASCPGKISVPPVRNLHQSGFS 540
 QY 541 LSGSQADHIAARTAKIMAPPGGRSNTISLS 572
 Db 541 LSGSQADHIAARTAKIMAPPGGRSNTISLS 572
 RESULT 5
 JC5318
 dihydroxyrimidinase related protein 3 - human
 C:Species: Homo sapiens (man)
 C:Date: 01-May-1997 #sequence_revision 18-Jul-1997 #text_change 02-Sep-2000
 C:Accession: JC5318
 R:Hamajima, N.; Matsuda, K.; Sakata, S.; Tamaki, N.; Sasaki, M.; Nonaka, M.
 Gene 180, 157-163, 1996
 A:Title: A novel gene family defined by human dihydroxyrimidinase and three related p
 A:Reference number: JC5315; MUID:97128821
 A:Accession: JC5318
 A:Status: nucleic acid sequence not shown
 A:Molecule type: mRNA
 A:Residues: 1-570 <HAM>
 A:Cross-references: DDBJ:D78014; NID:g1330241; PIDN:BAAL1192.1; PID:g1330242
 A:Experimental source: fetal brain
 A:Comment: This protein is the homologue of rat 64K turned on after division protein
 C:Genetics:
 A:Gene: GDB:IDPYSL3; DRP-3
 A:Cross-references: GDB:5885804
 C:Superfamily: allantoinase; Bacillus dihydroorotase homology

Query Match 73.2%; Score 2186; DB 2; Length 570;
 Best Local Similarity 69.9%; Pred. No. 7e-153;
 Matches 400; Conservative 91; Mismatches 79; Indels 2; Gaps 2;

QY 1 MSFGKKSIPRITSDRLIRGGRIYVNDQSFYADVHVHEDGLIKQIGENLIVPGGIHIDA 60
 Db 1 MSFGKKSIPRITSDRLIRGGRIYVNDQSFYADVHVHEDGLIKQIGENLIVPGGIHIDA 60
 QY 61 HGLMVLPGGVVHTRLOMPVLGTFPADDPCQGTAAALAGGTTMILDHVPFDTGVSLLAAY 120
 Db 61 HGLMVLPGGVVHTRLOMPVLGTFPADDPCQGTAAALAGGTTMILDHVPFDTGVSLLAAY 120
 QY 61 NGRWVPGGIDVHTFQMPYKGTMTVDFFQGTAAALAGGTTMILDHVPFDTGVSLLAAY 120
 Db 61 NGRWVPGGIDVHTFQMPYKGTMTVDFFQGTAAALAGGTTMILDHVPFDTGVSLLAAY 120
 QY 121 EQRERADSAACDYSILHVDITRWHSIKEELEALVKEGVNSFLVFMAYKDRCCSDSQ 180
 Db 121 EQRERADSAACDYSILHVDITRWHSIKEELEALVKEGVNSFLVFMAYKDRCCSDSQ 180
 QY 121 EQRERADSAACDYSILHVDITRWHSIKEELEALVKEGVNSFLVFMAYKDRCCSDSQ 180
 Db 121 EQRERADSAACDYSILHVDITRWHSIKEELEALVKEGVNSFLVFMAYKDRCCSDSQ 180
 QY 181 MYEFTSIRDLGALQAQVHAENGDIIVEEQKRLLELITGPEGHVLSPHEVEAEAVYRAV 240
 Db 181 MYEFTSIRDLGALQAQVHAENGDIIVEEQKRLLELITGPEGHVLSPHEVEAEAVYRAV 240
 QY 181 MYEFTSIRDLGALQAQVHAENGDIIVEEQKRLLELITGPEGHVLSPHEVEAEAVYRAV 240
 Db 181 MYEFTSIRDLGALQAQVHAENGDIIVEEQKRLLELITGPEGHVLSPHEVEAEAVYRAV 240
 QY 241 TIAQANCPVLYTVKMSKAADLISQARKKGNVVFEGEPIITASLGTDGSHYWSKNWAKAA 300
 Db 241 TIAQANCPVLYTVKMSKAADLISQARKKGNVVFEGEPIITASLGTDGSHYWSKNWAKAA 300
 QY 241 TIAQANCPVLYTVKMSKAADLISQARKKGNVVFEGEPIITASLGTDGSHYWSKNWAKAA 300
 Db 241 TIAQANCPVLYTVKMSKAADLISQARKKGNVVFEGEPIITASLGTDGSHYWSKNWAKAA 300
 QY 301 FWTSPVNPDPDTADHITCLSSGDLQVTSACHTFTTAQKAVGKDNFALIPGSTNGIEE 360
 Db 301 FWTSPVNPDPDTADHITCLSSGDLQVTSACHTFTTAQKAVGKDNFALIPGSTNGIEE 360
 QY 301 FWTSPVNPDPDTADHITCLSSGDLQVTSACHTFTTAQKAVGKDNFALIPGSTNGIEE 360
 Db 301 FWTSPVNPDPDTADHITCLSSGDLQVTSACHTFTTAQKAVGKDNFALIPGSTNGIEE 360
 QY 361 RMSVWKECVASGKMDENEFVAVTSTNAKIFNFPYPRKRVAVGSDADLVINPKATKII 420
 Db 361 RMSVWKECVASGKMDENEFVAVTSTNAKIFNFPYPRKRVAVGSDADLVINPKATKII 420
 QY 421 SAKTHNLNVEYNIFEGVPCRCAPAVVISQGRVALEDGKMFVTPGAGRFVPRKTPFPDFVYK 480
 Db 421 SAKTHNLNVEYNIFEGVPCRCAPAVVISQGRVALEDGKMFVTPGAGRFVPRKTPFPDFVYK 480
 QY 481 RIKARNRLAEIHGVPRGLYDGPVHEVMPAKPGCAPARASCPGKISVPPVRNLHQSGFS 540
 Db 481 RIKARNRLAEIHGVPRGLYDGPVHEVMPAKPGCAPARASCPGKISVPPVRNLHQSGFS 540

QY 301 FWTSPVNPDPDTADHITCLSSGDLQVTSACHTFTTAQKAVGKDNFALIPGSTNGIEE 360
 Db 301 FWTSPVNPDPDTADHITCLSSGDLQVTSACHTFTTAQKAVGKDNFALIPGSTNGIEE 360
 QY 361 RMSVWKECVASGKMDENEFVAVTSTNAKIFNFPYPRKRVAVGSDADLVINPKATKII 420
 Db 361 RMSVWKECVASGKMDENEFVAVTSTNAKIFNFPYPRKRVAVGSDADLVINPKATKII 420
 QY 421 SAKTHNLNVEYNIFEGVPCRCAPAVVISQGRVALEDGKMFVTPGAGRFVPRKTPFPDFVYK 480
 Db 421 SAKTHNLNVEYNIFEGVPCRCAPAVVISQGRVALEDGKMFVTPGAGRFVPRKTPFPDFVYK 480
 QY 481 RIKARNRLAEIHGVPRGLYDGPVHEVMPAKPGCAPARASCPGKISVPPVRNLHQSGFS 540
 Db 481 RIKARNRLAEIHGVPRGLYDGPVHEVMPAKPGCAPARASCPGKISVPPVRNLHQSGFS 540
 QY 541 LSGSQADHIAARTAKIMAPPGGRSNTISLS 571
 Db 541 LSGSQADHIAARTAKIMAPPGGRSNTISLS 571
 RESULT 4
 S55525
 Ulip protein - mouse
 C:Species: Mus musculus (house mouse)
 C:Date: 01-Aug-1995 #sequence_revision 01-Sep-1995 #text_change 02-Sep-2000
 C:Accession: S55525
 R:Byk, T.; Dobransky, T.; Cifuentes-Diaz, C.; Sobel, A.
 submitted to the EMBL Data Library, June 1995
 A:Description: Identification and molecular characterisation of Ulip, a putative mammal
 A:Reference number: S55525
 A:Accession: S55525
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-570 <BYK>
 A:Cross-references: EMBL:X87817; NID:g861058; PIDN:CAA61082.1; PID:g861059
 C:Genetics:
 A:Gene: Ulip
 C:Superfamily: allantoinase; Bacillus dihydroorotase homology

Query Match 73.3%; Score 2187; DB 2; Length 570;
 Best Local Similarity 70.1%; Pred. No. 5.9e-153;
 Matches 401; Conservative 90; Mismatches 79; Indels 2; Gaps 2;

QY 1 MSFGKKSIPRITSDRLIRGGRIYVNDQSFYADVHVHEDGLIKQIGENLIVPGGIHIDA 60
 Db 1 MSFGKKSIPRITSDRLIRGGRIYVNDQSFYADVHVHEDGLIKQIGENLIVPGGIHIDA 60
 QY 61 HGLMVLPGGVVHTRLOMPVLGTFPADDPCQGTAAALAGGTTMILDHVPFDTGVSLLAAY 120
 Db 61 HGLMVLPGGVVHTRLOMPVLGTFPADDPCQGTAAALAGGTTMILDHVPFDTGVSLLAAY 120
 QY 61 NGRWVPGGIDVHTFQMPYKGTMTVDFFQGTAAALAGGTTMILDHVPFDTGVSLLAAY 120
 Db 61 NGRWVPGGIDVHTFQMPYKGTMTVDFFQGTAAALAGGTTMILDHVPFDTGVSLLAAY 120
 QY 121 EQRERADSAACDYSILHVDITRWHSIKEELEALVKEGVNSFLVFMAYKDRCCSDSQ 180
 Db 121 EQRERADSAACDYSILHVDITRWHSIKEELEALVKEGVNSFLVFMAYKDRCCSDSQ 180
 QY 121 EQRERADSAACDYSILHVDITRWHSIKEELEALVKEGVNSFLVFMAYKDRCCSDSQ 180
 Db 121 EQRERADSAACDYSILHVDITRWHSIKEELEALVKEGVNSFLVFMAYKDRCCSDSQ 180
 QY 181 MYEFTSIRDLGALQAQVHAENGDIIVEEQKRLLELITGPEGHVLSPHEVEAEAVYRAV 240
 Db 181 MYEFTSIRDLGALQAQVHAENGDIIVEEQKRLLELITGPEGHVLSPHEVEAEAVYRAV 240
 QY 181 MYEFTSIRDLGALQAQVHAENGDIIVEEQKRLLELITGPEGHVLSPHEVEAEAVYRAV 240
 Db 181 MYEFTSIRDLGALQAQVHAENGDIIVEEQKRLLELITGPEGHVLSPHEVEAEAVYRAV 240
 QY 241 TIAQANCPVLYTVKMSKAADLISQARKKGNVVFEGEPIITASLGTDGSHYWSKNWAKAA 300
 Db 241 TIAQANCPVLYTVKMSKAADLISQARKKGNVVFEGEPIITASLGTDGSHYWSKNWAKAA 300
 QY 241 TIAQANCPVLYTVKMSKAADLISQARKKGNVVFEGEPIITASLGTDGSHYWSKNWAKAA 300
 Db 241 TIAQANCPVLYTVKMSKAADLISQARKKGNVVFEGEPIITASLGTDGSHYWSKNWAKAA 300
 QY 301 FWTSPVNPDPDTADHITCLSSGDLQVTSACHTFTTAQKAVGKDNFALIPGSTNGIEE 360
 Db 301 FWTSPVNPDPDTADHITCLSSGDLQVTSACHTFTTAQKAVGKDNFALIPGSTNGIEE 360
 QY 301 FWTSPVNPDPDTADHITCLSSGDLQVTSACHTFTTAQKAVGKDNFALIPGSTNGIEE 360
 Db 301 FWTSPVNPDPDTADHITCLSSGDLQVTSACHTFTTAQKAVGKDNFALIPGSTNGIEE 360
 QY 361 RMSVWKECVASGKMDENEFVAVTSTNAKIFNFPYPRKRVAVGSDADLVINPKATKII 420
 Db 361 RMSVWKECVASGKMDENEFVAVTSTNAKIFNFPYPRKRVAVGSDADLVINPKATKII 420
 QY 421 SAKTHNLNVEYNIFEGVPCRCAPAVVISQGRVALEDGKMFVTPGAGRFVPRKTPFPDFVYK 480
 Db 421 SAKTHNLNVEYNIFEGVPCRCAPAVVISQGRVALEDGKMFVTPGAGRFVPRKTPFPDFVYK 480
 QY 481 RIKARNRLAEIHGVPRGLYDGPVHEVMPAKPGCAPARASCPGKISVPPVRNLHQSGFS 540
 Db 481 RIKARNRLAEIHGVPRGLYDGPVHEVMPAKPGCAPARASCPGKISVPPVRNLHQSGFS 540

Db 481 RIKARRKADHAYPRGMYDGPVFDLTPKGTGPGSARGSTRPN-PPVRNLHQSFGS 539
 Qy 541 LSGSQADHIAARRTAQKIMAPPGGRSNTLS 572
 Db 540 LSGTQVDEGV-RSASKRIVAPPGRSNTLS 570

RESULT 6
 JC5316
 dihydropyrimidinase related protein 1 - human
 C:Species: Homo sapiens (man)
 C:Date: 01-May-1997 #sequence_revision 18-Jul-1997 #text_change 01-Dec-2000
 C:Accession: JC5316; S58890
 R:Hamajima, N.; Matsuda, K.; Sakata, S.; Tamaki, N.; Sasaki, M.; Nonaka, M.
 Gene 180, 157-163, 1996
 A:Title: A novel gene family defined by human dihydropyrimidinase and three related proteins
 A:Reference number: JC5316; MUID:97128821
 A:Status: nucleic acid sequence not shown
 A:Molecule type: mRNA
 A:Residues: 1-572 <HAM>
 A:Cross-references: DBJ:U78012; NID:g1330237; PIDN:BA11190.1; PID:g1330238
 A:Experimental source: fetal brain
 R:Goshima, Y.; Nakamura, F.; Strittmatter, P.; Strittmatter, S.M.
 Nature 376, 509-514, 1995
 A:Title: Collapsin-induced growth cone collapse mediated by an intracellular protein related to dihydropyrimidinase
 A:Reference number: S58889; MUID:95364923
 A:Accession: S58890
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: mRNA
 A:Residues: 64-503, 'H', 505-572 <GOS>
 A:Cross-references: EMBL:U17278; NID:g882148; PID:AA93201.1; PID:g882149
 A:Note: The nucleotide sequence was submitted to the EMBL Data Library, November 1994
 C:Comment: This protein is involved in nervous system development.
 C:Genetics:
 A:Gene: GDB:CRMP1; DRP-1; DPYSL1
 A:Cross-references: GDB:5585714
 C:Superfamily: allantoinase; Bacillus dihydroorotase homology

Query Match 72.1%; Score 2151; DB 2; Length 572;
 Best Local Similarity 69.0%; Pred. No. 2.7e-150;
 Matches 394; Conservative 85; Mismatches 92; Indels 0; Gaps 0;

Qy 1 MSFGKSPRITSRLIRGGRIYNDQSFYADVHEDGLIKQIGENLIVPGTIDA 60
 Db 1 MSYGGKSPHITSRLIRGGRIYNDQSFYADVHEDGLIKQIGENLIVPGTIDA 60
 Qy 61 HGLMVLPGVDVHTRQLQMPVLGTPADFCQGTAAALAGGTTMILDHVFPDTSVLLAAY 120
 Db 61 NGRMVLPGVDVHTRQLQMPVLGTPADFCQGTAAALAGGTTMILDHVFPDTSVLLAAY 120
 Qy 121 EQWRERASACCDYSLHVDITRWHESTKEEALVKEGVNSFLVFMAYKDRQCSDSQ 180
 Db 121 EKWHAEADKSCCDYSLHVDITRWHESTKEEALVKEGVNSFLVFMAYKDRQCSDSQ 180
 Qy 181 MYEFTSLRDLGALQAQVHAENGDIIVEEQKRLLELGTGPGHVLSPHEEAEAVYRAV 240
 Db 181 LYEATFRLGALQAQVHAENGDIIVEEQKRLLELGTGPGHVLSPHEEAEAVYRAV 240
 Qy 241 TIAQANCPYVTKVMSKGAADATAQAKRRGVVFGPITASLGTGSHYWSKNWAKAAA 300
 Db 241 TIAGRINCPVITVMSKGAADATAQAKRRGVVFGPITASLGTGSHYWSKNWAKAAA 300
 Qy 301 FVTSPPVNPDPPTADHLTCLSSGDLQVTSAGHCTFTTAQKAVGKONFALIPETNGIEE 360
 Db 301 FVTSPPVNPDPPTADHLTCLSSGDLQVTSAGHCTFTTAQKAVGKONFALIPETNGIEE 360
 Qy 361 RMSMWEKCVASGKMDENEFVAVTSTNAKIFNFPYPRKGVAVGSDADLVINWPKATKII 420
 Db 361 RMTVWVDKAVATGKMDENEFVAVTSTNAKIFNFPYPRKGVAVGSDADLVINWPKATKII 420

Qy 421 SAKTHNLNVEYIFEGVECRGAPAVVISQGRVALEDGKMFETPGAGRFVPRKTFDFVVK 480
 Db 421 TAKSHKSAVEYIFEGMECHGSPVVISQGKIVFEDGINVKNKGRIKPAFEHLIQ 480
 Qy 481 RIKARNRLAEIHGVPRGLYDGVHEVMVPAKPGSCAPARASCPCGKISVPPVNRNLHQSFGS 540
 Db 481 RVKIRNKVFGQLQVSRGMYDGPVEVPATPKYATPAPSAKSSPSKHPPTIRNLHQSNGS 540

RESULT 7
 JC5315
 dihydropyrimidinase (EC 3.5.2.2) - human
 N:Alternate names: 5,6-dihydropyrimidine amidohydrolase; Hydantoinase
 C:Species: Homo sapiens (man)
 C:Date: 01-May-1997 #sequence_revision 18-Jul-1997 #text_change 18-Aug-2000
 C:Accession: JC5315
 R:Hamajima, N.; Matsuda, K.; Sakata, S.; Tamaki, N.; Sasaki, M.; Nonaka, M.
 Gene 180, 157-163, 1996
 A:Title: A novel gene family defined by human dihydropyrimidinase and three related proteins
 A:Reference number: JC5315; MUID:97128821
 A:Accession: JC5315
 A:Status: nucleic acid sequence not shown
 A:Molecule type: mRNA
 A:Residues: 1-519 <HAM>
 A:Cross-references: DBJ:U78011; NID:g2339965; PIDN:BA11189.1; PID:g1330236
 A:Experimental source: liver
 C:Comment: This enzyme is Zn2+-metalloenzyme, and the second enzyme involved in uracil
 dihydrolysis to N-carbamyl-beta-aminoisobutyrate. It catalyzes the hydrolysis of a vari
 A:Genetics:
 A:Gene: GDB:DPYS; DHPase
 A:Cross-references: GDB:5885803
 C:Complex: homotetramer
 C:Superfamily: allantoinase; Bacillus dihydroorotase homology
 C:Keywords: hydrolase

Query Match 52.1%; Score 1554; DB 2; Length 519;
 Best Local Similarity 59.3%; Pred. No. 1.9e-106;
 Matches 292; Conservative 69; Mismatches 127; Indels 4; Gaps 1;

Qy 16 RLLIRGGRIYNDQSFYADVHEDGLIKQIGENLIVPG----GIHTDAHGLMVLPGVD 71
 Db 6 RLLIRGGRIYNDQSFYADVHEDGLIKQIGENLIVPG----GIHTDAHGLMVLPGVD 65
 Qy 72 VHTRLQMPVLGTPADFCQGTAAALAGGTTMILDHVFPDTSVLLAAYEOWRERASAA 131
 Db 66 THTHMQFPFGMSRSDIDFHQGTAAALAGGTTMILDHVFPDTSVLLAAYEOWRERASAA 125
 Qy 132 CCDSYSLHVDITRWHESTKEEALVKEGVNSFLVFMAYKDRQCSDSQMYEFTSIRDL 191
 Db 126 CCDSYSLHVDITRWHESTKEEALVKEGVNSFLVFMAYKDRQCSDSQMYEFTSIRDL 185
 Qy 192 GALAQVHAENGDIIVEEQKRLLELGTGPGHVLSPHEEAEAVYRAVYRAVYRAV 251
 Db 186 GALAQVHAENGDIIVEEQKRLLELGTGPGHVLSPHEEAEAVYRAVYRAVYRAV 245
 Qy 252 VTKVMSKGAADATAQAKRRGVVFGPITASLGTGSHYWSKNWAKAAAFTVSPVNPDP 311
 Db 246 IVHVMKSAKAVTADARRDGRGVYGEPIAASLGTGTHYWNKEHHAHVHVGPPRLPDP 305
 Qy 312 TTADHLTCLSSGDLQVTSAGHCTFTTAQKAVGKONFALIPETNGIEERMSWKEKVA 371
 Db 306 STPDFLMNLANDLTTCTDNCFTNTCOKALGKODFTKIPNGVNGVEDRMSWKEKVA 365
 Qy 372 SKMDENEFVAVTSTNAKIFNFPYPRKGVAVGSDADLVINWPKATKIIISAKTHNLNVEY 431
 Db 366 SKMDENEFVAVTSTNAKIFNFPYPRKGVAVGSDADLVINWPKATKIIISAKTHNLNVEY 425
 Qy 432 NIFEGVECRGAPAVVISQGRVALEDGKMFETPGAGRFVPRKTFDFVVKIRKARNRLAEI 491

RESULT 9
T20007
hypothetical protein c47E12.8 - Caenorhabditis elegans
C.Species: Caenorhabditis elegans
C.Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Aug-2000
C.Accession: T20007

RESULT 10
T23968
hypothetical protein R05C7.3 - *Caenorhabditis elegans*
C:Species: *Caenorhabditis elegans*
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 02-Sep-2000
C:Accession: T23968
R:Gardner, A.
submitted to the EMBL Data Library, April 1996
A:Reference number: Z19825
A:Accession: T23968
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA

RESULT 10
T23968

A:Residues: 1-544 <WIL>
A:Cross-references: EMBL:Z71266; PIDN:CAA95842.1; GSPDB:GN00019; CESP:R06C7.3
A:Experimental source: clone R06C7
C:Genetics:
A:Gene: CESP:R06C7.3
A:Map position: 1
A:Introns: 103/3; 135/2; 195/3; 255/3; 325/3; 418/2; 440/2; 522/3
C:Superfamily: allantoinase; Bacillus dihydroorotase homology

Query Match 37.6%; Score 1122; DB 2; Length 544;
Best Local Similarity 46.7%; Pred. No. 1.2e-74;
Matches 218; Conservative 79; Mismatches 136; Indels 34; Gaps 4;

QY 14 SDRLLRGRIVNDQSFYADV-----HVEDGLIKOI 45
Db 81 SPLVLKNGTVNEDGMEKADLVNTVDVSLQLISCNENFFLNTWQYTKLHRH---REV 137
QY 46 GENLIVPGGIHTIDAHGLMVLPGVDVHTRLQMPVLGMPADDFCQGTKAALAGGTTMIL 105
Db 138 SPNITALPOTEVIDATRLVIFGGIDPHFMQPMYMGVETKDDFLKGTAAVAGGTTMII 197
QY 106 DHVFPD--TGVSLAAYEQWRERADSACCDYSLHVDITRWHESEKEELEALVK--EKGVN 162
Db 198 DFCCPDHNGESLIAGYNRWSWADPKVCCDYGLSVATMRPETAQMAITSPFEGVN 257
QY 163 SFLVFMAYKDRQCSDSOMYEIIFSIIRDLGALAQVHAENGDIIVEEQKRLLELGTITGPEG 222
Db 258 SFKFMAYENTLWVRDDELFRQMCALRALARVHCENSGSVIKEKEIDLLAKGVGTPEG 317
QY 223 HVLSPHEVEAEAVRYAVTIAKQANCLPYVTKVMSKGAADAIAQAKRGVVVFGPEPTAS 282
Db 318 HTQSPREEIEAETNRACVLAQAACNCPYIVHVMKGAASAISHHRAQGSIVFGEPTAAG 377
QY 283 LGTDSHYSKWNWAKAAAFVTPPVNDPPTADHLTCLLSSGDLQVTSAGHCTFTTAQKA 342
Db 378 LALDGSYINEDWLHARYVMSPLSRDPTPELLMLLAAGELHGLTGDCYDCRQS 437
QY 343 VGKONFALIPGTTNGIEERMSVWKECVASGKMDENEFVAVTSTNAKIFNFPYPRKGRVA 402
Db 438 LGKGNFTKIPNGINGVDRMSVWKEGVHSGIIDPMRYVTSITSTAAKIFNFPYPRKGRIA 457
QY 403 VGSADLVVNNPKATKIISAKTNLNVYINFEFVCGRGAPAVVISO 449
Db 498 VGSADLVVNNPKATKIISAKTNLNVYINFEFVCGRGAPAVVISO 544

RESULT 11
H83590
dihydropyrimidinase PA0441 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C>Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: H83590
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; B.
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
.; Lory, S.; Olson, M.V.
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
Nature 406, 959-964, 2000
A:Reference number: AB2950; MUID:2043737
A:Accession: H83590
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-479 <STO>
A:Cross-references: GB:AE004481; GB:AE004091; NID:99946293; PIDN:AA03830.1; GSPDB:GN001
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA0441
C:Superfamily: allantoinase; Bacillus dihydroorotase homology

Query Match 36.0%; Score 1074.5; DB 2; Length 479;
Best Local Similarity 45.8%; Pred. No. 3.1e-71;
Matches 216; Conservative 81; Mismatches 166; Indels 9; Gaps 4;

QY 17 LLIRGRIVNDQSFYADVHVEDGLIKQIGENLIVPGGIHTIDAHGLMVLPGVDVHTRL 76
Db 3 LLIRGATVVTHEESYRADVLCAAGLQIAIGENLETPSGCDVLDGGQYVLPGGIDPHTHM 62
QY 77 QMPVLGMPADDFCQGTKAALAGGTTMILHVDVPTDGTGVSLLAAYEQWRERADSACCDYS 136
Db 63 QLPFWGTVASEDFFGTAAGLAGGTTSIIDFIPNPROSLLEAFHTWRGWAQSA-ADYG 121
QY 137 LHVDTIRWHESEKEELEALVKQGVNSFLVFMAYKDRQCSDSOMYEIIFSIIRDLGALAQ 196
Db 122 FHVATVWNSDEVAREMGEVLAHQGVNSFKHFWAYKATMAAADDTLVASFERCLGAVPT 181
QY 197 VHAENGDIIVEEQKRLLELGTITGPEGHVLSHPEVEAEAVRYAVTIAKQANCLPYVTKVM 256
Db 182 VHAENGELVLFHQKLLAQGLTGPFAHPLSRPPOVEGEASRAIRIAETLGTPLYLVIHS 241
QY 257 SKGAADAIAQAKRGVVVFGPEPTASLTGDSHYSKWNWAKAAAFVTPPVNDPPTADH 316
Db 242 SREALDEITAVARAKGQPVYGEVLAGHLLDDSVYRHPDMATAAGYVNSPPFRP----VEH 297
QY 317 LTCL---LSSGDLQVTSAGHCTFTTAQKAVGKDNFALIPGTTNGIEERMSVWKECVASG 373
Db 298 QEALWRGLQSGNLHTTATDHCFCACQKAMGRDDFSKIPNGTAGIEDRMALLDAGVNSG 357
QY 374 KMDENEFVAVTSTNAKIFNFPYPRKGRVAVGSDADLVVNNPKATKIISAKTNLNVYINI 433
Db 358 RLSMHFEVALTSTNTAKIFNFPYPRKGRVAVGSDADLVVNNPKATKIISAKTNLNVYINI 433
QY 434 FEGVECGRPAVVISGRVALEDGKMFVTPGAGRFYPRKTFDFVYVYKRIKAR 485
Db 418 FEGTVRGIPSHSTISQGLLWAAAGDLRBPAGRYVERPAYDS-VYEVLGRR 468

RESULT 12
JC2310
dihydropyrimidinase (EC 3.5.2.2) - Bacillus stearothermophilus
N:Alternate names: hydantoinase
C:Species: Bacillus stearothermophilus
C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: JC2310; PC2206
R:Mukohara, Y.; Ishikawa, T.; Watabe, K.; Nakamura, H.
Biosci. Biotechnol. Biochem. 58, 1621-1626, 1994
A:Title: A thermostable hydantoinase of Bacillus stearothermophilus NS1122A: Cloning,
A:Reference number: JC2310; MUID:95036870
A:Contents: NS1122A
A:Accession: JC2310
A:Molecule type: DNA
A:Residues: 1-471 <MUK>
A:Cross-references: GB:S73773; NID:9688287; PIDN:AA060487.1; PID:9688288
A:Accession: PC2206
A:Molecule type: protein
A:Residues: 1-20 <M02>
C:Comment: This enzyme is a metalloenzyme and the oligomeric structure is required fo
C:Superfamily: allantoinase; Bacillus dihydroorotase homology
C:Keywords: hydrolase

Query Match 28.0%; Score 837; DB 1; Length 471;
Best Local Similarity 40.0%; Pred. No. 8.7e-54;
Matches 183; Conservative 79; Mismatches 192; Indels 4; Gaps 4;

QY 18 LIRGRIVNDQSFYADVHVEDGLIKQIGENLIVPGGIHTIDAHGLMVLPGVDVHTRLQ 77
Db 4 LIRKNGTIVTADTIYEADLLIQDKIAVIGRNL-DESGAEVIDATGCVYFPGGIDPHTHLD 62
QY 78 MPVLGMPADDFCQGTKAALAGGTTMILHVDVPTDGTGVSLLAAYEQWRERADSACCDYS 137
Db 63 MPFGTVTKDDFESGTIAAAGGTTTIIDFCLTNKGEPLKKAETWNNKATGKAVIDYGF 122
QY 138 HVDITRWHESEKEELEALVKQGVNSFLVFMAYKDRQCSDSOMYEIIFSIIRDLGALAQ 197
Db 123 HLMISEITDDVLEELPKVIEEGTISFKVMAYKDVQFQADDGTLYRTLVAAKELGALVM 182

Db 610 DVPLRG--APDRMIGALSTQPLVVTCTSHRPVNSATRVAAKD-FAIAQKSGTGAERMA 666
 QY 364 MWKEKVASCKMDENEFVAVTSTNAKIFENFYPRKGRVAVGSDADLVINWPKATKIIIAK 423
 Db 667 VVHERAVSRGRIIDAMRFVAVTSTNAKMFNPKKGRVAVGADADLVINWASGKRVLESS 726
 QY 424 THNINVEYNIFEGVEGRGAPVVISQGRVALEDGKMFVTPGAGRFVPRKTFDFVYKRIK 483
 Db 727 RAOSSQSNMVDGLTVHSVVTATVIGKIAIYONGEVRAPVAGFLRLSPNPFVLSMVG 786
 QY 484 ARNRLAEIHGVPGCLYDGPVHEVMVPAKPGSGAPAPARASCPCGKISVPPVRLNH-----535
 Db 787 QRDKPFANVER-----EASSQQ-----KPOQNGHHKSGDFD 821
 QY 536 -----QSGFSLSGSQADDAHARTTAQKIMAPPGBS 566
 Db 822 RNRTKVMESIDFGS-----AANRPRNPPGGRT 850
 RESULT 14
 T28685
 hypothetical protein - Streptomyces coelicolor
 C:Species: Streptomyces coelicolor
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 20-Jun-2000
 C:Accession: T28685
 R:Parkhill, J.; Bentley, S.D.; Barrell, B.G.; Rajandream, M.A.
 submitted to The EMBL Data Library, January 1999
 A:Reference number: Z20512
 A:Accession: T28685
 A:Status: preliminary; translated from GB/EMBL/DDBJ
 A:Molecule type: DNA
 A:Residues: 1-467 <PAR>
 A:Cross-references: EMBL:AL023496; PIDN:CAA18902.1
 C:Superfamily: allantoinase; Bacillus dihydroorotase homology
 Query Match: 23.2%; Score 692.5; DB 2; Length 467;
 Best Local Similarity 33.8%; Pred. No. 3.7e-43;
 Matches 163; Conservative 84; Mismatches 192; Indels 43; Gaps 10;
 QY 14 SDRLLIRGRVNDQSFYADVHVEDGLIKOI-----GENLIVPGGIHTIDAH 61
 Db 2 SNTVIRGLVLTASDETHADVLEDGRVAALAAATGTPAAEAFTANV-----IDAS 53
 QY 62 GLMLVPGVDHTRLOMPVLGNTPADDFCOGTAAALAGTTMILDHVPDPTGVSLAAEY 121
 Db 54 GKYPVPGVDGTHMEMPFGGTAAATFETGTRAAANGTTTIVDFAIQSVHSLREGLD 113
 QY 122 QWRERADSAACDYSILHVDITRWHSIEKEELAVKGVNSFLVFMAYKDRCCSDSQM 181
 Db 114 AWHAKAEGNCAIDYGFHMIVSDVNQETLKEMDLLV-EEGVTSTFKQFMAYPGVYSDGQI 172
 QY 182 YEIITSIRDLGALAOVHAENGDIIVEEOKRLELGTGTPGEGHVLSPHEVEAEAVYRAVT 241
 Db 173 LEAQRAAENGGLIMHAENGTAIDVLVEQALARGETDPRPHGEVVKALLAEATHRAIR 232
 QY 242 IAKOANCLVTVKMSKGAADATAQAKRRGVVVFGE--PITASLGTDGSHYKSNWA---296
 Db 233 LAQVAGAPLYVHVHVSATEAVAELTRARDEGLPVFGTGPYLFSTLD-----NLAEPD 285
 QY 297 -KAAAFVTSPPVNDPTADHLTCL---LSSGDLQVTSAGHCTFT--TAQAVGKDNFALI 351
 Db 286 FEGAKYVCSTPLRP-----KEHQAAALWGLTNDLVQVSTHCPFCFSGOKELGRGDSRI 341
 QY 352 PEGTNGIERMSVWYKCVASGKMDENEFVAVTSTNAKIFENFYPRKGRVAVGSDADLV 411
 Db 342 PNGMPGVENRMDLL-HQAVVEGHIGRRRWRWIEIACATPARMFLTPKAGTAPACADIVV 400
 QY 412 WNPKATKIIIAKTHNLNVEYNIFEGVEGRGAPVVISQGRVALEDGKMFVTPGAGRFVPR 471
 Db 401 YDPHAEQVISAETHMNDYISAYEGRRTIGRVEIVLSRGEVPVVTREYTGKKGHAYTPR 460

QY 198 HAENGDIVBEEOKRLELGLITGPEGHVLSHPEVEAEAVYRAVTITAKOANCLVYTKVMS 257
 Db 183 HAENGDIVDLYTKKALEDGTDPIYHALTRPPELEGEATGRACQCLTELAGSGLYVHVSC 242
 QY 258 KGAADATAQAKRRGVVVFGEPIPTASIGTDSHWKSNWAKAAAFVTSPPVNDPTADHL 317
 Db 243 AQAVEKTAEARNGLNVMGTCTCPQYLVLSQSYLEKFN-EGAKYVMSPLR-EKWHQEVYL 300
 QY 318 TCLSSGDLQVTSAGHCTFT-TAQKAVGKDNFALIEPGTNGIERMSVWYKCVASGKMD 376
 Db 301 WNALKNQQLTQSGDQCSDFKQKELGRGDTKIPNGGPILIEDRVSLISFSEGVKGRIT 360
 QY 377 ENFVAVTSTNAKIFENFYPRKGRVAVGSDADLVINWPKATKIIIAKTHNLNVEYNIFEG 436
 Db 361 LNQFVDIVSTRIAKLGLFPKKGTTIAGADADLVIPDPTVERVISAETHMAMVDYNPFE 420
 QY 437 VECRCAPAWISQGRVALEDGKMFVTPGAGRFVPRKTF 474
 Db 421 MKVTGEPVSLRCGEFVYVRDQKQFVGKPGYGYVVRKRY 458
 RESULT 13
 S33558
 unc-33 protein - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 07-Apr-1994 #sequence_revision 02-Jun-1994 #text_change 21-Jul-2000
 C:Accession: S33558; S33559; S24643; S24644
 R:Li, W.; Herman, R.K.; Shaw, J.E.
 Genetics 132, 675-689, 1992
 A:Title: Analysis of the Caenorhabditis elegans axonal guidance and outgrowth gene unc-33
 A:Reference number: S33558; MUID:93106371
 A:Accession: S33558
 A:Molecule type: DNA
 A:Residues: 1-854 <LIW1>
 A:Cross-references: EMBL:Z14148; NID:g6899; PIDN:CAA78520.1; PID:g6900
 A:Accession: S33559
 A:Molecule type: mRNA
 A:Residues: 1-854 <LIW2>
 A:Cross-references: EMBL:Z14146; NID:g6903; PID:g6904
 C:Genetics:
 A:Introns: 14/2; 68/1; 129/3; 151/3; 182/1; 255/3; 361/2; 660/1; 790/3
 A:Keywords: alternative initiators
 F:1-854/Product: unc-33 protein (long form) #status predicted <MARI>
 F:176-854/Product: unc-33 protein (intermediate form) #status predicted <MARI>
 F:332-854/Product: unc-33 protein (short form) #status predicted <MARI>
 Query Match: 26.2%; Score 781.5; DB 2; Length 854;
 Best Local Similarity 32.4%; Pred. No. 2.6e-49;
 Matches 187; Conservative 114; Mismatches 204; Indels 73; Gaps 11;
 QY 5 GKKSIPRTSDRLIRGGRIVNDQSFYADVHVEDGLIKOIGENIIVPGGIHTIDAHGLM 64
 Db 330 GENSI-----LLVNAQIVNDDAIFAIVADILIEDGIIONVAPNLEAPEGAEVLDAAGKL 382
 QY 65 VLPQGVVDHTRLOMPVLGNTPADDFCOGTAAALAGTTMILDHVPDPTGVSLAAEYQWR 124
 Db 383 ALPAGIDVYTVT-----DSSVDLSTGCKSAIAGTGTIVEVVRPRAESVSAVKRVK 437
 QY 125 ERAD-SAACDYSILHVDITRWHSIEKEELAVKGVNSFLVFMAYKDRCCSDSQMVE 183
 Db 438 NOLEKSGISCHVALSVAITDFCE---QENSELYKNGINSFVL-----DGVSLTDDKLE 489
 QY 184 IFSIIRDLGALAOVHAENGDIIVEEOKRLELGTGTPGEGHVLSPHEVEAEAVYRAVTIA 243
 Db 490 LFEHVHKLALIRVVPENKSVAMLEKMKLKGVTGPEFPQRPRESLEADRVSGVCVLG 549
 QY 244 KOANCLVTVKMSKGAADATAQAKRRGVVVFGEPIPTASIGTDSHWKSNWAKAAAFV 303
 Db 550 NIASCPISIVQVSSADSALAEKARASALAHAEIASAAVTAADGSAFSDLRFAHAHLT 609
 QY 304 SPPVNDPTADHLTCLSSGDLQVTSAGHCTFTTAQKAVGKDNFALIEPGTNGIERMS 363

Search completed: July 30, 2001, 11:41:24
Job time: 238 sec


```

68 GlyClyValAspValHisThrArgLeuGlnMetProValLeuGlyMetTh 84
157 GCGGGTATTGATTCGACACACGATTTAGATATCCCGTTTGGCGCACGGT 206
84 rProAlaAspPheCysGlnGlyThrLysAlaAlaLeuAlaGlyGlyT 101
207 GACAAAGGATGATTTCGATCTGGACGATTTGCGCGGCATTTGGCGGAA 256
101 hrThrMetLeuAspHisValPheProAspThrGlyValSerLeuLeu 117
257 CAACGACCATCATCTGACTTTGTTTAAACGAATAAAGGGGACCATTA 306
118 AlaAlaThrGluLeuThrArgGluArgAlaAlaSerAlaAlaCysAs 134
307 AAAGCGATTGAACCTGGGCACACAAAGCGAAGGAAAGCGGTATTGA 356
134 pTyrSerLeuHisValAspLeuThrArgTrpHisGluSerLeuGlu 151
357 TTATGGCTTCCATTAAATGATTAGCGAATTAACGATGACGTATTAGA 406
151 LuLeuGluAlaLeuValLysGlyLysValAsnSerPheLeuValPhe 167
407 ACCTGCCAAAGATCATTCGCCGAAGGAGGATAACATCCTTTAAAGTGT 456
168 MetAlaTyrLysAspArgCysGlnCysSerAspSerGlnMetTyrGlu 184
457 ATGGCGTATAAAGCGTATTTCAGCGACGATGATGGAACGTATTATAC 506
184 ePheSerLeuLeuArgAspLeuGlyAlaLeuAlaGlnValHisAlaGlu 201
507 GCTAGTGCTGCCAAAGAACCTTGGCGCGCTGTGCTGTTTCATGCGGAA 556
201 snGlyAspLeuValGluGluGlnLysArgLeuLeuGluGlyLe 217
557 ATGGGATGCTGATTGATTAACGAAAGGAAAGCGCTTGGCGAAGGAAT 606
218 ThrGlyProGluGlyHisValLeuSerHisProGluGluValGluAla 234
607 ACGAGCGCGATTTACCATGCTTTAACGCGCCCTCCAGAGTAGAAGAGA 656
234 uAlaValTyrArgAlaValThrLeuAlaLysGlnAlaAsnCysProLeu 251
657 AGCAGCGCGCGCGCTGCTCAATTCAGAGAGCTTGGCGGTTCACAACTT 706
251 yrValThrLysValMetSerLysGlyAlaAlaAspAlaLeuAlaGln 267
707 ACCTTGTTCAGTGACATGCGCAAGCGGTGGAAATAATTCACCAAGCG 756
268 LysArgArgGlyValValPheGlyGluProIleThrAlaSerLeuGl 284
757 CGCAATAAGGGTGGATGTGTGGGAGAAACGTGCGCAATATCTTGT 806
284 yThrAspGlySerHisTyrTrpSerLysAsnTrpAlaLysAlaAlaP 301
807 TCTCGACCAATCG...TATTAGAAAGCGCTGATTTGAAGCGCGAAAT 853
301 heValThrSerProValAsnProAspProThrThrAlaAspHisLeu 317
854 ATGTTGGTCCCTCCGCTCGT...GAAATAATGCAATCAAGATATTG 900
318 ThrCysLeuLeuSerSerGlyAspLeuGlnValThrGlySerAlaHis 334
901 TGGATGCGCTGAAACAGCGCGCTGCAAGCGCTTGGATCGGACCAATG 950
334 sThrPheThr...ThrAlaGlnLysAlaValGlyLysAspAsnPheAla 350
951 TTCATTTCATTTAAAGGCCAAAGAAAGAACTTGGCAGAGAGATTTACTA 1000
350 euLeuProGluGlyThrAsnGlyIleGluGluArgMetSerMetValTrp 366
1001 AAATTCACAAAGCGCGCGATGTCGAGGATCGGTCAGCATCTTTTC 1050
367 GluLysCysValAlaSerGlyLysMetAspGluAsnGluPheValAla 383

```

```

1051 AGTGAAGGGTTAAAGAGGAAGATCACTTAATCAATTTGTCGATAT 1100
383 lThrSerThrAsnAlaAlaLysIlePheAsnPheTyrProArgLysGly 400
1101 TATGTCACAGAATTCGAATTTGTCGGGTATTCCCGAGAAAGGAA 1150
400 rgValAlaValGlySerAspAlaAspLeuValIleTrpAsnProLysAla 416
1151 CGATCGCGGTAGTTTCAGACGACGACTTAGTCAATTTGACCCGGATATC 1200
417 ThrLysIleLeuSerAlaLysThrHisAsnLeuAsnValGluTyrAsn 433
1201 GAACGGGTGATTTCCGGCGGAACACACATATGCGCGTCACTATAATGC 1250
433 ePheGluGlyValGluCysArgGlyAlaProAlaValValIleSerGln 450
1251 ATTTGAAGGAATGAAGTAACGGGTGAACCGGTATCGGTTCTGTGCAGAG 1300
450 lyArgValAlaLeuGluAspGlyLysMetPheValThrProGlyAlaGly 466
1301 GCGAATTTGTTGTCGCTGATAAACAATTTGTCGGAACACAGGTACGCG 1350
467 ArgPheValProArgLysThrPhe 474
1351 CAATATTAAAGCTGCGTGT 1374

```

seq_name: /cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-602-656-2

```

seq_documentation_block:
; Sequence 2, Application US/08602656
; Patent No. 5679571
; GENERAL INFORMATION:
; APPLICANT: Burtscher, Helmut; Lang, Gunter; Popp, Friedrich
; TITLE OF INVENTION: Recombinant D-Hydantoinase, A Process For the
; TITLE OF INVENTION: Production and Use
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; APPLICATION DATA:
; CURRENT APPLICATION NUMBER: US/08/602,656
; FILING DATE: 16-FEB-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/289,709
; FILING DATE: 12-AUGUST-1994
; APPLICATION NUMBER: P 43 28 829.4
; FILING DATE: 27-AUGUST-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, No. 5679571man D.
; REGISTRATION NUMBER: 30,946
; REFERENCE/DOCKET NUMBER: BOER 1041
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1383 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-602-656-2

```

us-09-367-496-8.std.rni

Tue Jul 31 13:08:19 2001

```

alignment_scores:
  Quality: 826.00      Length: 458
  Ratio: 2.614        Gaps: 4
  Percent Similarity: 68.996      Percent Identity: 39.738

alignment_block:
US-09-367-496-8 x US-08-602-656-2 ..

Align seg 1/1 to: US-08-602-656-2 from: 1 to: 1383

18 LeuileArgGlyGlyValGlyValAlaAspAspGlnSerPheValAlaAs 34
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
10 AFAATAAATAATGAAACATGTTACCGCAACCGATACGTATGAAGCGGA 59
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
34 pValHisValGluAspGlyLeuileLysGlnIleGlyGluAsnLeuilev 51
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
60 CTTGCTCATTAAGACGCGAAATTTGCCATGATAGGCCAACATTTA...G 106
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
51 alProGlyClyIleHisThrIleAspAlaHisGlyLeuMetValLeuPro 67
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
107 AAGAAAAAGCGCTGAAGTGTATGATGCAAGGCTGTACGTATTTCCA 156
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
68 GlyGlyValAspValHisThrArgLeuGlnMetProValLeuGlyMetTh 84
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
157 GCGGTTATGATTCGACACGCGATTTAGATATGCGGTTTGGCGGCACG 206
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
84 rProAlaAspPheCysGlnGlyThrIlyAlaAlaLeuAlaGlyGlyT 101
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
207 GACAAAGGATGATTTGCAATCTGGAACGATTCGCGCGGCATTTGGCG 256
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
101 hrThrMetIleLeuAspHisValProAspThrGlyValSerLeuLeu 117
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
257 CACGACCATCATCGACTTTGTTTAAACCAATAAAGGGGACCATTAATA 306
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
118 AlaAlaThrGluGlnTrpArgGluArgAlaAspSerAlaAlaCysCys 134
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
307 AAAGCGATTAAGAACTTGGCAACAAAGGAGGAAAGCGGTATTGGA 356
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
134 pThrSerLeuHisValAspIleThrArgTrpHisGluSerIleLysGlu 151
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
357 TTATGCTTCCATTTATGATGATAGCGAAATACGGATGACGTATTAGA 406
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
151 luLeuGluAlaLeuValLysGluLysGlyValAsnSerPheLeuValPhe 167
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
407 AGCTGCCAAAGATCATTTGCCCAAGAGGGGATAACATCTTTAAAGT 456
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
168 MetAlaThrLysAspArgCysGlnCysSerAspSerGlnMetThrGlu 184
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
457 ATGGCGTATAAAGACGTATTTCAGCGAGATGATGGAACGTATACCGC 506
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
184 ePheSerIleLeuArgAspLeuGlyAlaLeuAlaGlnValHisAlaGlu 201
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
507 GCTAGTGGCTGCCAAGAACTTGGCGGCTGTGTCATGTTTCATGCGGAA 556
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
201 snGlyAspIleValGluGluGlnLysArgLeuLeuGluLeuGlyLe 217
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
557 ATGGGGATGTGATGATTACTTAACGAAAGAAAGCGCTTGGGAAGGGAAT 606
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
218 ThrGlyProGluGlyHisValLeuSerHisProGluGluValGluAla 234
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
607 ACGGAGCGGATTTACCATGCTTTAAGCGGCTCCAGAAAGTAGAAGGAG 656
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
234 uAlaValThrArgAlaValThrIleAlaLysGlnAlaAsnCysProLeu 251
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
657 ACGACCGCGCGCGCTGTCAATTCAGACAGCTTCCCGGTTCAACAAT 706
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
251 yValThrLysValMetSerLysGlyAlaAlaAspAlaIleAlaGlnAla 267
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
707 ACCTGTTTCCACGTGACATGTCGCAAGCGGTGGAAGAAATTCACAACG 756
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
268 LysArgArgGlyValValValPheGlyGluProIleThrAlaSerLeu 284
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

757 CGCAATAAAGGGTGGATGTGGGGAGAAACGCTGCCCAATATCTTGT 806
284 yThrAspGlySerHisTrpTrpSerLysAsnTrpAlaLysAlaAlaAla 301
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
807 TCTCGACCAATCG...TATTTAGAAAAGCCTGATTTTGAAGCGCGAAAT 853
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
301 heValThrSerProProValAsnProAspProThrThrAlaAspHisLeu 317
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
854 ATGTTTGGTCCCTCCCTCGT...GAAAAATGCGCATCAAGAAGTATTG 900
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
318 ThrCysLeuLeuSerSerGlyAspLeuGlnValThrGlySerAlaHisCy 334
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
901 TGGATGCGCTGAAAAACCGCCAGCTGCAACGCTGGATCGGACCAATG 950
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
334 sThrPheThr...ThrAlaGlnLysAlaValGlyLysAspAsnPheAla 350
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
951 TTCAATTTGACTTTAAAGCGCAAAAGAACTTGGCAGAGGAGATTTACTA 1000
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
350 euIleProGluGlyThrAsnGlyIleGluGluArgMetSerMetValTrp 366
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1001 AAATTCCAACCGCGCGGCGATGGTCGAGGATCGGGTCAGCATTTCTTC 1050
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
367 GluLysCysValAlaSerGlyLysMetAspGluAsnGluPheValAla 383
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1051 AGTGAAGGGTTAAAGAAAGAAAGAACTACGTTAAATCAATTTGTCGAT 1100
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
383 lThrSerThrAsnAlaAlaLysIlePheAsnPheTrpProArgLysGly 400
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1101 TATGTCGACAAAGAAATGCGCAATTTGTCGGTTATTTCCGAGAAAAGAA 1150
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
400 rgValAlaValGlySerAspAlaAspLeuValIleTrpAsnProLysAla 416
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1151 CGATCGCGGTAGGTTTCAGACGCGACACTTAGTCATTTTACCCCGGAT 1200
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
417 ThrLysIleIleSerAlaLysThrHisAsnLeuAsnValIleLysAsn 1433
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1201 GAAGCGGTGATTTCCGCGGAAACACACCATATGCGCGTCGACTATAAT 1250
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
433 ePheGluGlyValGluCysArgGlyAlaProAlaValValIleSerGln 450
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1251 ATTTGAAGGAATGAAGTAACCGGTGAACCGGTATCGGTTCTGTGAGAG 1300
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
450 lyArgValAlaLeuGluAspGlyLysMetPheValThrProGlyAlaGly 466
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1301 GCGAATTTGTCGCGTGATTAACAATTTTCGCGAAAACCGAGGTACGGC 1350
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
467 ArgPheValProArgLysThrPhe 474
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1351 CAATATTTAAAGAGCTGGCTGTTT 1374

seq_name: /cgn2_6/ptodata/1/lna/58_COMB.seq:US-08-815-356-1

seq_documentation_block:
: Sequence 1, Application US/08815356
: Patent No. 5858759
: GENERAL INFORMATION:
: APPLICANT: Neal, Robert J
: APPLICANT: Griffin, Alison M
: APPLICANT: Scott, Miller O
: APPLICANT: Schatzman, Allan R
: APPLICANT: Gorham, Hazel C
: TITLE OF INVENTION: D-N-Carbamoyl-Amino Acid Amidohydrolase
: TITLE OF INVENTION: and Hydanolase
: NUMBER OF SEQUENCES: 2
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: SmithKline Beecham Corporation/Corporate
: ADDRESS: Intellectual Prop.
: STREET: P.O. Box 1539-UW2220
: CITY: King of Prussia
: STATE: Pennsylvania
: COUNTRY: USA
: ZIP: 19406-0939
: COMPUTER READABLE FORM:

```

```

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/815.356
; FILING DATE: 11-MAR-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/356.369
; FILING DATE: 23-DEC-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Jarvis, Herbert H
; REGISTRATION NUMBER: 31,171
; REFERENCE/DOCKET NUMBER: P30433
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (610) 270-5019
; TELEFAX: (610) 270-5090
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1880 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 891..1805
; US-08-815-356-1

```

alignment_scores:

```

Quality: 197.50 Length: 141
Ratio: 2.244 Gaps: 4
Percent Similarity: 62.411 Percent Identity: 39.716

```

alignment_block:

```

US-09-367-496-8 x US-08-815-356-1/rev ..

```

```

Align seg 1/1 to reverse of: US-08-815-356-1 from: 1 to: 1880

```

```

17 LeuLeuIleArgGlyGlyArgIleValAsnAspGlnSerPheTyrAl 33
   ::::::::::::::::::::|
419 ATCATCATCAAGAACGGAACCATCGTAACCGCGGACGGGATTCCTCCGC 370
   ::::::::::::::::::::|
33 AspValHisValIleAspGlyLeuIleGlyGlnIleGly.GluAsnLeu 49
   ::::::::::::::::::::|
369 CGATCTCGGAATCAAGATGCGAAGATCGCCAGATCGCGGGAACGTTCCG 320
   ::::::::::::::::::::|
50 IleValProGlyGlyLeuHisThrIleAspAlaHisGlyLeuMetValLe 66
   ::::::::::::::::::::|
319 GCCCGCGCGCGCGGA.....CAAATCGACGCTCCGCGCGCTACGTTT 276
   ::::::::::::::::::::|
66 uProGlyGlyValAspValHisThrArgLeuGlnMetProValLeuGlyM 83
   ::::::::::::::::::::|
275 TCGCGCGCATCGAGTTTCATCATGATCGAGACGTCAGCTTCAACA 226
   ::::::::::::::::::::|
83 etThrProAlaAspPheCysGlnGlyThr..LysAlaAlaLeuAlaG 99
   ::::::::::::::::::::|
225 CGCAGTCGCGCGCACATTCGCAACCGCGACGTCGCGCGCC...TGTG 179
   ::::::::::::::::::::|
99 lYgIyThrThrMetIleLeuAspHisValPheProAspThrGlyValSer 115
   ::::::::::::::::::::|
178 GCGGACGACGACCATCGTCGATTCCTGCCAGCAGGACCGCGGCATAGC 129
   ::::::::::::::::::::|
116 LeuLeuAlaAlaTyrGluGlnTrpArgGluArgAlaAspSerAlaAlaCy 132
   ::::::::::::::::::::|
128 CTGAGGAGGCGGTGCCAAATGGAGCGCATGCCCGCGCAAGTCGCG 79
   ::::::::::::::::::::|
132 scYsAspTyrSerLeuHisValAspIleThrArgTrpHisGluSerIleL 149
   ::::::::::::::::::::|
78 GATCGACTACGGCTACCATATCATCGTGTCTGTCGACTGATGACGTGA 29
   ::::::::::::::::::::|

```

```

149 ysGluGluLeuGluAlaLeu 155
   ::::::::::::::::::::|
28 TCGAGAGCTAGAGTACTG 9
   ::::::::::::::::::::|
seq_name: /cgn2_6/ptodata/1/ina/5A_COMB.seq:US-07-732-242C-8
seq_documentation_block:
; Sequence 8, Application US/07732242C
; Patent No. 5298399
; GENERAL INFORMATION:
; APPLICANT: Uozumi,Takeshi; Masaki,Haruhiko;
; APPLICANT: Hidaka,Makoto; Nakamura,Akira;
; APPLICANT: Maeda,Michihsa; Yoneta,Yasuo
; TITLE OF INVENTION: Gene of Urease
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Frishauf,Holtz,Goodman & Woodward,P.C.
; STREET: 600 Third Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10016-2088
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 0.72mb
; COMPUTER: IBM PC compatible (NEC PC-9801 RX)
; OPERATING SYSTEM: MS DOS
; SOFTWARE: ASCII Form
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07732,242C
; FILING DATE: 19910718
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JPN 2-210178
; FILING DATE: 10-AUG-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Goodman, Herbert
; REGISTRATION NUMBER: 17081
; REFERENCE/DOCKET NUMBER: 910532/HG
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)972-1400
; TELEFAX: (212)370-1622
; TELEX: 236268
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6131 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: both
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-07-732-242C-8

```

alignment_scores:

```

Quality: 196.00 Length: 616
Ratio: 0.737 Gaps: 26
Percent Similarity: 43.182 Percent Identity: 19.805

```

alignment_block:

```

US-09-367-496-8 x US-07-732-242C-8 ..

```

```

Align seg 1/1 to: US-07-732-242C-8 from: 1 to: 6131

```

```

1 MetSerPheGlnGlyLysLysSerIle.....Pr 10
   ::::::::::|
1106 GTAAAGTTTGGCGCGCAAGGTGATCCGAGATGGAATGGGCGCATCC 1155
   ::::::::::|
10 oArgIleThrSerAspArg.....LeuLeuIleArgGlyGlyArgI 24
   ::::::::::|
1156 TTGGCGCAACAGGATGAATGCGTTCATCTCGTATTAAACAATGGCATTA 1205
   ::::::::::|
24 leValAsnAspAspGlnSerPheTyrAlaAspValHisValGluAspGly 40
   ::::::::::|
1206 TTGTTGATTACACAGGTATTATTATAAAGCAGATATCGGCATAAAGATGA 1255
   ::::::::::|

```

tue Jul 31 13:08:19 2001

```

41  LeuIleLysGlnIleGlyGlu..... 47
1256 ATGATTGCTCCATAGAAAGCGGGAACCCCTTCTTAATGACGGGT 1305
48  AsnLeuIleValProGlyGlyIleHisThrIleAspAlaHisGlyLeuM 64
1306 CGATATGGTGTGGACGACGACAGCAAGTCATAGCCGAGAGGATGA 1355
64  etValLeuProGlyGlyValAspValHisThrArgLeuGlnMetProVal 80
1356 TTGTGACAGCGGAGGAGATAGATCTCATATTCATTTATTTGCCCT... 1402
81  LeuGlyMetThrProAlaAspPheCysGlnClyThrLysAlaAlaLe 97
1403 .....CAGCAATCGAACCCTCT 1422
97  uAlaGlyGly...ThrThrMetIleLeuAspHisValPheProAspThrG 113
1423 TGCATCGGGTGTGACCATATGATGGCGGAGGAACACAGGACCCGCTACAG 1472
113  lyValSerLeuLeuAlaIleTyrGluGlnIleTyrArgGluArgAlaAspSer 129
1473 GCACA.....AAT 1480
130  AlaAlaCysCysAspTyrSerLeuHisValAspIleThrArgTrpHisG 146
1481 GCCACTACTGT.....ACACCGGGCGCTTG 1506
146  uSerIleLysGluGluLeuAlaLeuValLysGluLysGlyValAsnS 163
1507 GAATATCCATCGTATGCTTCAAGCA.....GCCGAAG 1538
163  erPheLeuValPheMetAlaTyrLysAspArgCysGlnCysSerAspSer 179
1539 AATTCGGATAACTTGGGCTTTTAGGAAAGGGAACATGTTACAT... 1585
180  GlnMetTyrGluIlePheSerIleLeuArgAspLeuGlyAlaLeuAlaG 196
1585 ..... 1585
196  nValHisAlaGluAsnGlyAspIleValGluGluGlnLysArgLeuL 213
1586 .....GAGGCTCCTTTAAGGACAAA 1607
213  euGluLeuGlyIleThrGlyProGluGlyHisValLeuSerHisProGlu 229
1608 TTGAACGGGCGGTGGGATTAAGCTTAC.....GAA 1642
230  Glu.....ValGluAlaGluAlaValTyrArgAlaValThrIleAlaL 244
1643 GATTGGGATCGACGGCGGCTTATTGATACATGTTTGAAGTGGCGGA 1692
244  sGlnAlaAsnCysProLeuTyrVal...ThrLysValMetSerLysGly. 259
1693 TCGATATGATGCAAGTATGCGGATTCATACAGACACTTAAATGAAGCG 1742
260  ....AlaAlaAspAlaIleAlaGlnAlaLysArgArgGlyValValVal 274
1743 GATTTGTGAGGATACTTTGAAGCCCATAGACGGTGCAGTGATTCATACC 1792
275  PheGlyGluProIleThrAlaSerLeuGlyThrAspGlySerHisTyrTr 291
1793 TAT.....CATACAGAAGGGCTGGCGGGGACAT..... 1822
291  pSerLysAsnTrpAlaLysAlaAlaPheValThrSerProProVala 308
1823 .GCTCGGATATATAAAGCGCGGCTTCCGGAATATTGCTCTCTT 1871
308  snProAspPro.....ThrThrAlaAspHisLeu 317
1872 CCACGAATCCAACTCGACCTTATCTATCACTTTTGGAGAGCATTTA 1921
318  ThrCysLeuLeuSerSerGlyAspLeuGlnValThrGlySerAlaHisCy 334
1922 GATATGTTAATGTTGCCACCACTAGCGCTAATATTATTCAGAGGATAT 1971
334  sThrPheThrThrAlaGlnLysAlaValGlyLysAspAsn..... 347
1972 TGCCTTTT.....GCCATTCACCATACGAAAGAGAGACCATCGCGCGG 2015
348  ....PheAlaLeuIleProGluGlyThr 355
2016 AAGATGTTTTACATGATTTAGCGCTTTACAGCATGATTTCTGTGATTCA 2065
356  AsnGly.....IleGluGluArgMetSerMetValTrpGluLysCy 369
2066 CAGGCGATGGCGGAGTAGGAGAAGTATCTGCTAGTGGCAACGCG 2115
369  s.....ValAlaSerGlyLysMetAspGluAsn..... 378
2116 TGACAAGATGAAAGCAAGAGGAGGAGTTACAAGAACACAACTGCTGG 2165
379  ....GluPheValAlaValThrSerThrAsnAla 388
2166 GAGACAACCTTCGTGTGAACGTTATATTCGCAAAATATACGATCAATCCG 2215
389  AlaLysIlePheAsnPheTyrProArgLysGlyArgValAlaValGlySe 405
2216 GCCATTCGCTCATGTTATTCGGGATTTATGTTGGTGTGAAGTGGGAA 2265
405  rAspAlaAspLeuValIleTrpAsnProLysAlaThrLysIleIleSeRa 422
2266 ATTAGCTGATTTAGTGTGTGAATCCT..... 2293
422  lalYThrHisAsnLeuAsnValGluTyrAsnIlePheGluGlyValGlu 438
2294 .....GCTTTTTTGTGTGAA 2311
439  CysArgGlyAlaProAlaValIleSerGlnGlyArgValAlaLeuG 455
2312 .....CCTGAACCTGGTCTTAAAGAGGAGGATGTTGCTTACAG 2349
455  uAsp..... 456
2350 CACTATGGGAGATCCCAATGCCAGCATTCGCACACCGCGGTTTAT 2399
457  .GlyLysMetPheValThrProGlyAlaGlyArg..... 467
2400 ATCTGCTCGATGTTTGCAGCGAAGGAGATGCCAAATATCAACGCTCTATC 2449
468  ...PheValProArgLysThrPheProAspPheValTyrLysArgIleLy 483
2450 ACCTTTGTTCGAAAGCAGCGTATGAAAGGCGATTCATGAACAGTTGGG 2499
483  sAlaArgAsnArgLeuAlaGluIleHisGlyVal..... 494
2500 TTGGAAGAAAAGGTGAACACCGATCCATGGAATTCGAAATTCACGAAA 2549
495  ....ProArgGlyLeuTyrAspGlyPro 502
2550 AAGATTTAATTTTGAAACCATAAACCCCAAAATTCAGTCGATCCTCAG 2599
503  ValHisGluValMetValProAlaLysProGlySerGlyAlaProAla 518
2600 ACATATGAAGTAAAGTAGAGCGGTCAATTAGTAGCATGTGAACCGGCA 2647
seq_name: /cgn_2/ptodata/1/ina/5A_COMB.seq.us-08-967-513-1
seq_documentation_block:
; Sequence 1, Application 05/08967513
; Patent No. 5783436
; GENERAL INFORMATION:
; APPLICANT: Robert P. Hausinger
; TITLE OF INVENTION: Mutant Urease and Method
; TITLE OF INVENTION: Of Use For Determination

```

```

; TITLE OF INVENTION:  of Urea
; NUMBER OF SEQUENCES:  5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE:  Ian C. McLeod
; STREET:  2190 Commons Parkway
; CITY:  Okemos
; STATE:  Michigan
; COUNTRY:  USA
; ZIP:  48864
; COMPUTER READABLE FORM:
; MEDIUM TYPE:  Diskette, 5.25 inch, 360 KB
; MEDIUM TYPE:  storage
; COMPUTER:  IBM Compatible
; OPERATING SYSTEM:  MS-DOS
; SOFTWARE:  Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER:  US/08/967-513
; FILING DATE:  11-NOV-1997
; CLASSIFICATION:  435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:  08/587,645
; FILING DATE:  July 26, 1996
; ATTORNEY/AGENT INFORMATION:
; NAME:  Ian C. McLeod
; REGISTRATION NUMBER:  20,931
; REFERENCE/DOCKET NUMBER:  MSU 4.1-309
; TELECOMMUNICATION INFORMATION:
; TELEPHONE:  (517) 347-4100
; TELEFAX:  (517) 347-4103
; TELEX:  No. 5783436e
; INFORMATION FOR SEQ ID NO:  1:
; SEQUENCE CHARACTERISTICS:
; LENGTH:  2400
; TYPE:  nucleotides
; STRANDEDNESS:  Single
; TOPOLOGY:  Linear
; MOLECULE TYPE:
; DESCRIPTION:  cDNA
; HYPOTHEetical:  No
; ANTI-SENSE:  No
; ORIGINAL SOURCE:
; ORGANISM:  Klebsiella aerogenes
; STRAIN:  CG253
; INDIVIDUAL ISOLATE:
; CELL TYPE:  N/A
; FEATURE:
; NAME/KEY:  cDNA encoding mutant urease
; NAME/KEY:  h2190
; LOCATION:  Modification at position 1312 to
; LOCATION:  glutamine
; IDENTIFICATION METHOD:  Sequencing
; OTHER INFORMATION:
; US-08-967-513-1

alignment_scores:
      Quality:  142.00          Length:  583
      Ratio:    0.587           Gaps:    24
Percent Similarity:  41.509     Percent Identity:  20.926

alignment_block:
US-09-367-496-8 x US-08-967-513-1 ..

Align seg 1/1 to: US-08-967-513-1 from: 1 to: 2400

17 LeuLeutleArgGlyGlyArglleValAsnAspClnSerPheTyral 33
|||||: : : : : |||||: :
857 CTGGTGCTACCAACAGCGTGTATCGTCGACTACTGGGGATCGTTAAGC 906
|||||: : : : : |||||: :

33 aspValHisValGluaSpLyLeutleLysGlnlleGlyGlu..... 47
|||||: |||||: |||||: |||||:
907 CGATATCGCGGTGAAGCAGCCGCCGGATCTTCGCATCGGCAAGCGGCA 956

```

us-09-367-496-8.std.rni

Tue Jul 31 13:08:19 2001

```

1626 CGGACATCGCCGAGGACGTCGCTTTCGCGAGTCGCGCATTCGCCCGGAA 1675
    :
    :
    :
337 ThrThrAlaGlnLysAlaValGlyLysAsp.....AsnPheAlaLeu11 351
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1676 ACCATCGCTCGGAAAGACGTGCTGCAGCATCTCGGCGCTTCCTCGCTCAC 1725
    :
    :
    :
351 eProGluGlyThrAsnGly.....IleGluGluArgMetSerMetV 365
    :
    :
    :
1726 CTCCTCCGATTCGCAGGCCATCGGCGCGTCGCGGAGTAGTTCCTCGCA 1775
    :
    :
    :
365 alTrpGluLysCys.....ValAlaSerGlyLysMetAspGlu 377
    :
    :
    :
1776 CCTGCGAGGTGGCGCATCGCATGAAGGTGCAGCGCGGAGCGCTGGCGGAG 1825
    :
    :
    :
378 .....AsnGluPheValAlaValTh 384
    :
    :
    :
1826 GAGACCGGGGATACGACAACTTCGCGGTGAAGCGCTACATCGCACAATA 1875
    :
    :
    :
384 rSerThrAsnAlaLysIlePheAsnPheTyrProArgLysGlyArgV 401
    :
    :
    :
1876 CACCATCAACCGCGGCTGACCCACGGCATCGCACACGAAGTCGATCCA 1925
    :
    :
    :
401 aLalaValGlySerAspAlaAspLeuValIleTrpAsnProLysAlaThr 417
    :
    :
    :
1926 TTGAGGTGGTGAAGTGGCTGACCTCGTGGTCTGTGTACCA..... 1966
    :
    :
    :
418 LysIleIleSerAlaLysThrHisAsnLeuAsnValGluTyrAsnIlePh 434
    :
    :
    :
1967 .....GCCTT 1971
    :
    :
    :
434 eGluGlyValGluCysArgGlyAlaProAlaValValIleSerGlnGlyA 451
    :
    :
    :
1972 CTTCCGCGTGA.....CGGCCACCGTGTATCAAGCGGCA 2009
    :
    :
    :
451 rgValAlaLeuGluAsp..... 456
    :
    :
    :
2010 TGATCGCCATCGCGCGGATGGCGATATCAATGCCTCTATTTCGACCCCG 2059
    :
    :
    :
457 .....GlyLysMetPheValThrProGlyAlaGlyArgPh 468
    :
    :
    :
2060 CAGCCGGTGCACACTACCGCGGATGTTGGCGCGCTGGCGAGCGCGCCCA 2109
    :
    :
    :
468 eValProArgLysThrPhe.....ProAspPheValT 479
    :
    :
    :
2110 TCACCTGCGGCTCACCTTCCTCTCGCAGGCGCGGCAATGCGGCTTG 2159
    :
    :
    :
479 yrLysArgIleLysAlaArgAsnArgLeuAlaGluIleHisGlyValPro 495
    :
    :
    :
2160 CCGAGCGGCTGAACCTCGCAGCGCGATCGCGTGGTGAAGGCTGCCGT 2209
    :
    :
    :
496 ArgGlyLeuTyrAspGlyProValHisGluValMetValProAlaLysP 512
    :
    :
    :
2210 ACGGTGAGAAAGCGCATGTTGCACACAGTCTGCAGCGCTAACATCAC 2259
    :
    :
    :
512 roGlySerGlyAlaProAlaArgAlaSerCysProGlyLysIleSer 527
    :
    :
    :
2260 CGTGCAGCGCCAGACCTATGAGGTGCGGGTGATGGCGAACTTAICA 2306
    :
    :
    :
seq_name: /cgn2_6/ptodata/1/ina/5B_comb.seq:us-08-687-645B-1

seq_documentation_block:
; Sequence 1, Application US/08687645B
; Patent No. 5846752
; GENERAL INFORMATION:
; APPLICANT: Robert P. Hausinger
; TITLE OF INVENTION: Mutant Urease and Method
; TITLE OF INVENTION: of Use For Determination
; TITLE OF INVENTION: of Urea
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ian C. McLeod
; STREET: 2190 Commons Parkway

alignment_scores:
Quality: 142.00 Length: 583
Ratio: 0.587 Gaps: 24
Percent Similarity: 41.509 Percent Identity: 20.926

alignment_block:
US-09-367-496-8 x US-08-687-645B-1 ..
Align seg 1/1 to: US-08-687-645B-1 from: 1 to: 2400

17 LeuLeuIleArgGlyGlyArgIleValAsnAspGlnSerPheTyrAl 33
|||||:|||||:|||||:
857 CTGGTGTCTACCAACGCGTTCGTCGATCACTGGGGGATCGTTAAGGC 906
|||||:|||||:|||||:
33 aAspValHisValGluAspGlyLeuIleLysGlnIleGlyGlu..... 47
|||||:|||||:|||||:
907 CGATATCGGCGTGAAGACGCGCGGATCTTCGCCATCGGCAACCGCGCA 956
|||||:|||||:|||||:
48 .....AsnLeuIleValProGlyGlyIleHisThr... 57
|||||:|||||:|||||:
957 ACCCGGACATCCAGCCCAACGTCACCATCCCATCGGCGCTGCACGAA 1006
|||||:|||||:|||||:
58 ...IleAspAlaHisGlyLeuMetValLeuProGlyGlyValAspValH1 73
|||||:|||||:|||||:

```

```

1676 ACCATCGTGGGGAAGACGTGCTGCAGCATCTCGCGCCCTTCGCTCAC 1725
351 eProGluGlyThrAsnGly.....IleGluAluArgMetSerMetV 365
    :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
1726 CTCCTCCATTCGCGAGCCATGGCGCGCTGGGGAAGATGATTCTCGCA 1775
365 alTrpGluLysCys.....ValAlaSerGlyLysMetAspGlu 377
    :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
1776 CCTGCAGGTGGCGCATCGCATGAAGTGCAGCGCGGAGCGCTGGCGGAG 1825
378 .....AsnGluPheValAlaValAlaValTh 384
1826 GAGACCGGGGATACGACAACTTCGCGTGAAGCGCTACATCGCAATA 1875
384 rSerThrAsnAlaLysIlePheAsnPheTyProArgLysGlyArgV 401
    :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
1876 CACCATCAACCGCGCTGACCCACCGCATCGCACGAAGTCGGATCCA 1925
401 alAlaValGlySerAspAlaAspLeuValIleTrpAsnProLysAlaThr 417
    :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
1926 TTGAGTGGGTAACTGGCTGACCTCGTGTCTGTCAACA..... 1966
418 LysIleSerAlaLysThrHisAsnLeuAsnValGluTyAsnIlePh 434
1967 .....GCCTT 1971
434 eGluGlyValGluCysArgGlyAlaProAlaValIleSerGlnGlyA 451
    :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
1972 CTTGCGGGTGA.....CCGCCACCGTGATCAAAAGCGCA 2009
451 rgValAlaLeuGluAsp..... 456
    :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
2010 TGATGCCCATCGCGCGATGGCGATATCAATGCTCTATTTCGACCCCG 2059
457 .....GlyLysMetPheValThrProGlyAlaGlyArgPh 468
2060 CAGCCGGTGCACCTACCGCCCGATGTTGGCGCGTGGCGACGCCGCCCA 2109
468 eValProArgLysThrPhe.....ProAspPheValT 479
    :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
2110 TCACCTGCGCCTCACCTTCCTGTCGAGCGCGCGACCAATGGCGTTG 2159
479 yLysArgIleLysAlaArgAsnArgLeuAlaGluIleHisGlyValPro 495
    :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
2160 CCGAGCGCTGAACCTGCGCAGCGCATGCGCGTGGTGAAGGCTCCGT 2209
496 ArgGlyLeuTyAspGlyProValHisGluValMetValProAlaLysP 512
2210 ACGTGCAGAAAGCGACATGGTGCAACAAGCTCGACCTAACATCATC 2259
512 roGlySerGlyAlaProAlaArgAlaSerCysProGlyLysIleSer 527
    :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
2260 CGTGACGCCACGACCTATGAGTGGCGGTGGATGCGCACTATCA 2306

seq_name: /cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-467-822-19
seq_documentation_block:
seq: Sequence 19, Application US/08467822
Patent No. 5843460
GENERAL INFORMATION:
APPLICANT: Labigne, Agnes
APPLICANT: Sauerbaum, Sebastien
APPLICANT: Ferrero, Richard L.
APPLICANT: Thiberge, Jean-Michel
TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS AGAINST
TITLE OF INVENTION: HELICOBACTER INFECTION, POLYPEPTIDES FOR
TITLE OF INVENTION: COMPOSITIONS, AND NUCLEIC ACID SEQUENCES
TITLE OF INVENTION: POLYPEPTIDES
NUMBER OF SEQUENCES: 44
CORRESPONDENCE ADDRESS:
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
ADDRESSEE: Dunner
STREET: 1300 I Street, N.W.

```


Tue Jul 31 13:08:19 2001

CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/467,822
FILING DATE: 06-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/447,177
FILING DATE: 19-MAY-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/432,697
FILING DATE: 02-MAY-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Meyers, Kenneth J.
REGISTRATION NUMBER: 25,146
REFERENCE/DOCKET NUMBER: 03495.0137-02000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 408-4000
TELEFAX: (202) 408-4400
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 2619 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: misc_feature
LOCATION: 31..36
OTHER INFORMATION: /standard_name= "Shine-Dalgarno"
OTHER INFORMATION: sequence."
FEATURE:
NAME/KEY: misc_feature
LOCATION: 756..759
OTHER INFORMATION: /standard_name= "Shine-Dalgarno"
OTHER INFORMATION: sequence."
US-08-467-822-19

alignment_scores:
Quality: 136.50 Length: 553
Ratio: 0.576 Gaps: 24
Percent Similarity: 42.857 Percent Identity: 19.711

alignment_block:
US-09-367-496-8 x US-08-467-822-19 ..

Align seg 1/1 to: US-08-467-822-19 from: 1 to: 2619

1 MetSerPheGlnGlyLysSerile..... 9
852 ATCAAAATTTGGGGGGGTAATAAATATCGTGATGGGATGAGTCAAAACCA 941
10ProArgIleThrSerAspArgLeuLeuIleArgGlyGlyLeu 25
942 TAGCCCTAGCTCTTATGAAATAGATTGGTGTCTACTAACGCCCTCATG 991
25 alAsnAspGlnSerPheTyrAlaAspValHisValGluAspGlyLeu 41
992 TGGACTATACGGCCATTTACAAAGCCGACATTGGGATTAAAGACGGCAAG 1041
42 IleLysGlnIleGlyClu..... 47
1042 ATTGAGGCAATTGGCAAGCGGCAATAGGACATGAGATGGCGGTAGA 1091

48AsnLeuIleValProGlyGlyIleHisThrIleAspAlaHisGlyL 63
1092 TAATAATCTTTGGTAGGTCCTGCTACAGAGGCTTTGGCAGCTGAGGCT 1141
63 euMetValLeuProGlyGlyValAspValHisThrArgLeuGlnMetPro 79
1142 TGATTGTAACCGCTGGTGGCATCGATACGCAATATTCACCTTATCTCCC 1191
80 ValLeuGlyMetThrProAlaAspPheCysGlnGlyThrLysAlaAl 96
1192CAACAAATCCCTACTGCG 1208
96 alEuAlaGlyGly...ThrThrMetIleLeuAspHisValPheProAsp 112
1209 TTTTGGCAGCGGGTTACAAACCATGATTGGAGGAGCACAGGACCTCGG 1258
112 hrGlyValSerLeuLeuAlaAlaTyrGluGlnTrpArgGluArgAlaAsp 128
1259 ATGGCAGCAATGCG..... 1272
129 SerAlaAlaCysCysAspTyrSerLeuHisValAspIleThrArgTrpHi 145
1273ACCACCATCACTCCCGGACG 1292
145 sGluSerIleLysGluGluLeuAlaLeuValLysGlyVala 162
1293 CGTAATCTAAAGATGTTGTCGGTGA...GCCGAAGATACGCCATGA 1339
162 snSerPheLeuValPheMetAlaTyrLysAspArgCysGlnCysSerAsp 178
1340 AT.....CTAGGCTTTTGGCT...AAGGGGAATGTGCTTACGAACCC 1380
179 SerGlnMetTyrGluIle.....PheSerIleIleArg 189
1381 TCTTTACGGCATCAGATTGAACGAGCGGGGATTGTTTTAAATCCACGA 1430
189 gAspLeuGlyAlaLeu...AlaGlnValHisAlaGluAsnGlyAspIleV 205
1431 ACAGCTGGGGAAGCACACCTGCAGCTATTAC..... 1461
205 alGluGluGlnLysArgLeuLeuGluGlyIleThrGlyProGlu 221
1461 1461
222 GlyHisValLeuSerHisProGluValGluAlaGluAlaValTyrAr 238
1462 ...CACTGCCTCAATGTCGCGCATGAATACGATGTGCAAGTGGCTATCCA 1508
238 gAlaValThrIleAlaLysGlnAlaAsnCysProLeuTyrValThrLysV 255
1509 CACCGATACCTT...AACGAGCGGGCTGT..... 1536
255 alMetSerLysGlyAlaAlaAspAlaIleAlaGlnAlaLysArgArgGly 271
1537GTAGAAGACACCTAGAGCGGATTGCGGGCGCACCC 1572
272 ValValValPheGlyGluProIleThrAlaSerLeuGlyThrAspGlySe 288
1573 ATCCATACCTTC.....CACACTGAAGGGGCTGGGGTGG 1607
288 rHisTyrTrpSerLysAsnTrpAlaLysAlaAlaPheValThrSerP 305
1608 ACAC.....GCTCCAGATGTTATCAAAATGGCAGGGGAATTTAAACATTC 1651
305 roProValAsnProAspPro.....ThrThrAla 314
1652 TACCGCGCTCTACTAACCAGCCACCATCTCTTTCACCAAAACACACTGAAGCC 1701
315 AspHisLeuThrCysLeuLeuSerSerGlyAspLeuGlnValThrGlySe 331
1702 GAGCATGAGGACATGTTAATGCTGTGCCACCACCTTGGATAAAAGACTCAA 1751

```

331 rAlaHisCysThrPheThrThra.....GlnLysAlaValG 344
1752 GGAAGATGTGCGATTGCGGATTCGAGGATTCGCCCAAACTATCGCG 1801
344 lYlYsAspAsn.....PheAlaLeuIleProGluGly 354
1802 CTGAAGACCAACTCCATGATGATGGGATCTTTTATCATCACCAGCTCCGAC 1851
355 ThrAsnGly.....IleGluGluArgMetSerMetValTpsGlu 368
1852 TCTCAGGCTATGGAGCGGTAGCGGAGGTATCATCAGCAGCTTGCAGAC 1901
368 sCysValAlaSer.....GlyLysMetAspGluAsnGlu.... 379
1902 ACCAGCAAAAAACAAAAAGAGTTGGCGCTTGAAGAGGAAAAAGCG 1951
380 .....PheValAlaValThrSerThrAsn 387
1952 ATAAGCAGCAACTTCGCCATCAACCGCTACATCTCTAATACACCATCAAC 2001
388 AlaAlaLysIlePheAsnPheTyrProArgLysGlyArgValAlaValG 404
2002 CCCGGATCGCGCATGGGATTTCTGACTATGTGGCTCTGTGGAAGTGG 2051
404 yserAspAlaAspLeuValIleTpsAsnProLysAlaThrLysIleIles 421
2052 CAATACCGCCACCTGCTGCTTGGAGTCCG..... 2082
421 erAlaLysThrHisAsnLeuValGluTyrAsnIlePheGluGlyVal 437
2083 .....GCTTCTTGGCAT 2097
438 GlucYargGlyAlaProAlaValIleSerGlnGlyArgValAlaLe 454
2098 AAG.....CCCAATATGATTATTAAAGGCGGATTTATTGCGCT 2135
454 uGluAsp..... 456
2136 CTCTCAAAATGGCGATGCCAATCGCTATTCACCCCTCAGCCCGTCT 2185
457 .....GlyLysMetPheValThrProGlyAlaGlyArgPheValProArg 471
2186 ATTACCGTGAATGTTGGACACCATGGGAAAAACAAATTCGACACCAAT 2235
472 LysThrPhe 474
2236 ATCATTTT 2244

```

seq_name: /cgn2_6/ptodata/1/ina/6B_COMB.seq:US-08-432-697-19

```

seq_documentation_block:
; Sequence 19, Application US/08432697
; Patent No. 6248330
; GENERAL INFORMATION:
; APPLICANT: Labigne, Agnes
; APPLICANT: Sauerbaum, Sebastien
; APPLICANT: Ferrero, Richard L.
; APPLICANT: Thiberge, Jean-Michel
; TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS AGAINST
; TITLE OF INVENTION: HELICOBACTER INFECTION, POLYPEPTIDES FOR USE IN THE
; TITLE OF INVENTION: COMPOSITIONS, AND NUCLEIC ACID SEQUENCES ENCODING SAID
; NUMBER OF INVENTION: POLYPEPTIDES
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; ADDRESSEE: Dunner
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

```

```

; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/432,697
; FILING DATE: 02-MAY-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Meyers, Kenneth J.
; REGISTRATION NUMBER: 25,146
; REFERENCE/DOCKET NUMBER: 03495.0137-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 408-4000
; TELEFAX: (202) 408-4400
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2619 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 31..36
; OTHER INFORMATION: /standard_name= "Shine-Dalgarno
; OTHER INFORMATION: sequence."
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 756..759
; OTHER INFORMATION: /standard_name= "Shine-Dalgarno
; OTHER INFORMATION: sequence."
; US-08-432-697-19

```

alignment_scores:

```

Quality: 136.50 Length: 553
Ratio: 0.576 Gaps: 24
Percent Similarity: 42.857 Percent Identity: 19.711

```

alignment_block:

US-09-367-496-8 x US-08-432-697-19 ..

Align seq 1/1 to: US-08-432-697-19 from: 1 to: 2619

```

1 MetSerPheGlnGlyLysLysSerIle..... 9
892 ATCAAAATTGGGCGGTAAACCTATCCGTATGGGATGAGTCAACCAA 941
10 ....ProArgIleThrSerArgLeuLeuIleArgGlyGlyArgIleV 25
942 TAGCCCTAGCTCTTATGAATTAGATTGGTGTCTACCTAACGCCCTCAT 991
25 alAsnAspAspGlnSerPheTyrAlaAspValHisValGluAspGlyLeu 41
992 TGGACTATACGGCATTTACAAAGCCGACATTTGGATTAAAGACGCGCA 1041
42 IleLysGlnIleGlyGlu..... 47
1042 ATTGCAGGCATTGGCAAGCGACGCAATAAGGACATGCAAGATGGCGTAG 1091
48 ....AsnLeuIleValProGlyGlyIleHisThrIleAspAlaHisGly 63
1092 TAATAATCTTTGGCTAGTCTCTACAGAGGCTTTGGCAGCTGAGGGCT 1141
63 euMetValLeuProGlyGlyValAspValHisThrArgLeuGlnMetPro 79
1142 TGATTGTACCGCTGGTGGCATCGATACCATATTCACCTTTATCTCTCC 1191
80 ValLeuGlyMetThrProAlaAspPheCysGlnGlyThrLysAlaAl 96
1192 .....CAACAAATCCCTACTGC 1208
96 alLeuAlaGlyGly...ThrThrMetIleLeuAspHisValPheProAsp 112

```

Tue Jul 31 13:08:19 2001

```
1209 TTTTGCAGGGGTTTACAACTGATTTGGAGGAGGCACAGGAGCTGGCG 1259
112 hrGlyValSerLeuLeuAlaAlaTyrgluTrpArgGluArgAlaasp 128
1259 ATGGCAGGAATGGG..... 1272
129 SerAlaAlaCysCysAspTySerLeuHisValAspIleThrArgTrpHi 145
1273 ..... 1272
145 sGluSerIleLysGluLeuAlaLeuValLysGluLysGlyVala 162
1293 CGCTAATCTAAAGTAGTGTGGTGCA...GCCGAAGAAATACGCCATCA 1339
162 snSerPheLeuValPheMetAlaTyryLysAspArgCysGlnCysSerAsp 178
1340 AT.....CTAGGCTTTTGGCT...AAGGGAATGTGTCTTACGAACCC 1380
179 SerGlnMetTyrgluLeu.....PheSerIleIleAr 189
1381 TCTTTACGCGATCAGATTGAAGCAGGCGGATGTTTAAATCCACGA 1430
189 gasPheGlyAlaLeu...AlaGlnValHisAlaGluAsnGlyAspIleV 205
1431 AGACTGGGAGGACACACCTCGAGCTATTCAC..... 1461
205 aIcGluGluGlnLysArgLeuLeuGluGlyIleThrGlyProGlu 221
1461 ..... 1461
222 GlyHisValLeuSerHisProGluGluValGluAlaGluAlaValTrAr 238
1462 ...CACTGCTCATGTGCCGATGAATACGATGTGCAAGTGGTATCCA 1508
238 gAlaValThrIleAlaLysGlnAlaAsnCysProLeuTyValThrLysV 255
1509 CACCGATACCTT...AAGCAGCGGGGTGT..... 1536
255 alMetSerLysGlyAlaAlaAspAlaIleAlaGlnAlaLysArgGly 271
1537 .....GTAGAGACACCTAGAGCGGATTCGGCGGCACCC 1572
272 ValValValPheGlyGluProIleThrAlaSerLeuGlyThrAspGlySe 288
1573 ATCATACCTTC.....CACACTGAAGGGGTGGGGTGG 1607
288 rHisTyTrpSerLysAsnTrpAlaLysAlaAlaPheValThrSerP 305
1608 ACAC.....GCTCCAGATCTTATCAAAATGGCAGGGGAATTAACATTC 1651
305 roProValAsnProAspPro.....ThrThrAla 314
1652 TACCGGCTCTACTAACCCGACCATTCCTTTCACCAAAACACTGAAGCC 1701
315 AspHisLeuThrCysLeuSerSerGlyAspLeuGlnValThrGlySe 331
1702 GAGCAGATGGACATGTAATGGTGGCCACCACCTTGGATAAAGATCA 1751
331 rAlaHisCysThrPheThrAla.....GlnLysAlaValG 344
1752 GGAAGATGTCAGTTGGCGATTCGAGGATTCGCCCAAACTATCGCGG 1801
344 lLysAspAsn.....PheAlaLeuIleProGluGly 354
1802 CTGAAGACCAACTCCATGACATGGGGATCTTTTCTATCACCAGCTCCGAC 1851
355 ThrAsnGly.....IleGluGluArgMetSerMetValTrpGluLy 368
1852 TCTCAGGCTATGGACGCGTAGGCGGTGATCATCAGCAGCTTGGCAGAC 1901
368 sCysValAlaSer.....GlyLysMetAspGluAsnGlu..... 379
```

seq_name: /cyn2_6/ptodata/1/ina/5B_COMB.seq:US-08-920-095-1

```
seq_documentation_block:
; Sequence 1, Application US/08920095
; Patent No. 5837240
; GENERAL INFORMATION:
; APPLICANT: Cynthia K. Lee et al.
; TITLE OF INVENTION: MULTIMERIC, RECOMBINANT UREASE VACCINE
; NUMBER OF INVENTIONS: 3
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/920,095
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/431,041
; FILING DATE: 28-APR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Clark, Paul T.
; REGISTRATION NUMBER: 30,162
; REFERENCE/DOCKET NUMBER: 06132/013001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2735 base pairs
; TYPE: nucleic acid
```

STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-920-095-1

alignment_scores:

Quality: 136.50 Length: 632
Ratio: 0.493 Gaps: 29
Percent Similarity: 43.829 Percent Identity: 19.146

alignment_block:

US-09-367-496-8 x US-08-920-095-1 ..

Align seg 1/1 to: US-08-920-095-1 from: 1 to: 2735

```

17 LeuLeuIleArgGlyGlyArgIleValAlaAsnAspGlnSerPheTyrAl 33
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1023 TTAATTATCACTAACGCTTTAATCGTGGATTACACGGGTATTATATAAGC 1072
33 aspValHisValGluAspGlyLeuLeuLysGlnIleGlyGlu..... 47
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1073 GGATATTGTTAATTAAGATGGCAAAATCGCTGGCATTTGGTAAAGCGGTA 1122
48 .....AsnLeuIleValProGlyGly 54
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1123 ACAAGACATGCAAGATGCGTTAAACAACTTTAGCGTAGGTCCTGCT 1172
55 IleHisThrIleAspAlaHisGlyLeuMetValLeuProGlyGlyValas 71
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1173 ACTGAAGCCTTAGCGGTTGAAGTTTGTATCGTAACGGCTGGTGTATTGA 1222
71 pValHisThrArgLeuGlnMetProValLeuGlyMetThrProAlaAspA 88
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1223 CACACACATCCACTTCATTTCCACC..... 1247
88 spPheCysGlnGlyThrLysAlaAlaLeuAlaGlyGly...ThrThrMet 103
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1248 .....CAACAAATCCCTACAGCTTTTGAACGCGTGTAAACACCATG 1289
104 IleLeuAspHisValPheProAspThrGlyValSerLeuLeuAlaLay 120
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1290 ATTGGTGTGGAACCGTCTCTGTGTGAGGCACTAATGCG..... 1328
120 rGluGlnTrpArgGluArgAlaAspSerAlaAlaCysCysAspTyrSerL 137
1328 ..... 1328
137 euHisValAspIleThrArgTrpHisGluSerIleLysGluLeuGlu 153
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1329 .....ACTACTATCACTCCAGCGAGCAAGAAATTTAAATGGATGCTCAGA 1373
154 AlaLeuValLysGluLysGlyValAlaAsnSerPheLeuValPheMetAlay 170
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1374 GCG...GCTGAAGAATATCTATGANT..... 1397
170 rLysAspArgCysGlnCysSerAspSerGlnMetTyrGluIlePheSerI 187
1397 ..... 1397
187 leIleArgAspLeuGlyAlaLeuAlaGlnValHisAlaGluAsnGlyasp 203
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1398 .....TTAGTTTCTTGGCTAAGGTAAAGTAACTTCAACGATGCG 1436
204 IleValGluGluGlnLysArgLeuLeuGluLeuIleThrGly... 219
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1437 AGCTTAGCGGATCAA.....ATTGAAGCGGTGGGATTTGGCTT 1474
220 .....ProGlu.....GlyHisValL 225
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1475 TAAATTCACGAAGACTGGGCGACCACTCTCTTCGCAATCAATCATCGGT 1524
225 euSerHisProGluGluValGluAlaGluAlaValTyrArgAlaValThr 241
|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1525 TAGATGTTGGGACAAATACAGTGTCAAGTCGCTATCCACACACACT 1574
242 IleAlaLysGlnAlaAsnCysProLeuTyrValThrLysValMetSerLys 258
||:|||||:|||||:|||||:|||||:|||||:|||||:
1575 TTG...AATGAAGCGGTGT..... 1592
258 sGlyAlaAlaAspAlaIleAlaGlnAlaLysArgArgGlyValValValp 275
||:|||||:|||||:|||||:|||||:|||||:|||||:
1593 .....GTAGAAGACTATGGCTGCTATTGCTGGAGCGCACTATGCACACTT 1638
275 heGlyGluProIleThrAlaSerLeuGlyThrAspGlySerHisTyrTrp 291
||:|||||:|||||:|||||:|||||:|||||:|||||:
1639 TC.....CACACTGAAGCGGTGGCGCGGACAC..... 1667
292 SerLysAsnTrpAlaLysAlaAlaPheValThrSerProProValas 308
||:|||||:|||||:|||||:|||||:|||||:|||||:
1668 GCTCTGATATTATAAGTAGCGGTGAACACAAACATTCTTCCCGCTTC 1717
308 nProAspPro.....ThrThrAlaAspHisLeuT 318
||:|||||:|||||:|||||:|||||:|||||:|||||:
1718 CACTAACCCCCACCCTTTCACCGTGAATACAGAAGCAGACACATGG 1767
318 hrCysLeuLeuSerSerGlyAspLeuGlnValThrGlySerAlaHisCys 334
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1768 ACATGCTTATGGTGTGCCACCCTTGGATAAAAGCATTTAAAGAAGATGTT 1817
335 ThrPheThrThrAla.....GlnLysAlaValGlyLysAspAs 347
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1818 CAGTTGCTGATTCAAGGATCCGCCCTCAACCATTTGGCGGTGAAGACAC 1867
347 n.....PheAlaLeuIleProGluGlyThrAsnGly. 357
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1868 TTTGCATGACATGGGGATTTTCTCAATCACCAGTTCTGACTCTCAAGCA 1917
358 .....IleGluGluArgMetSerMetValTrpGluLysCysValAla 371
||:|||||:|||||:|||||:|||||:|||||:|||||:
1918 TGGCGCGTGTGGGTGAAGTATACACTAGAACCTTGGCAACACAGCTGACAA 1967
372 Ser.....GlyLysMetAspGluAsnGlu..... 379
||:|||||:|||||:|||||:|||||:|||||:|||||:
1968 AACAAAGAAAGAAATTTGGCGCTTGAAGAAGAAAGGCGGTAACGACAA 2017
380 .....PheValAlaValThrSerThrAsnAlaAlaLysI 391
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2018 CTTAGGATCAACCGCTACTTGTCTAAATACACCATTAAACCCAGCGATCG 2067
391 lePheAsnPheTyrProArgLysGlyArgValAlaValGlySerAspAla 407
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2068 CTCATGGGATTAGCGAGTATGTAGTTTCAGTAGAAGTGGGCAAGTGGCT 2117
408 AspLeuValIleTrpAsnProLysAlaThrLysIleleSerAlaLysTh 424
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2118 CACTGGTATTGTGGAGTCCA..... 2138
424 rHisAsnLeuAsnValGluTyrAsnIlePheGluGlyValGluCysArg 441
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2139 .....GCATTCTTTGGCGTGAA..... 2156
441 lyAlaProAlaValValIleSerGlnGlyArgValAlaLeuGluAspGly 457
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2157 .....CCCAACATGATCATCAAGCGGATTCATTGCGTTAAGCCAA... 2198
458 LysMetPheValThrProGlyAlaGlyArgPheValProArgLysThrPh 474
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2199 .....ATGGCGATGCGAAGCGTCTTATCCCTACC..... 2228
474 eProAspPheValTyrLysArgIle.....LysAlaA 485
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2229 .CCACACCGGTTTATACAGAGAAATGTTGCTCATCATGTTAAAGCTA 2277
485 rgAsnArgLeuAlaGluIleHisGlyValProArgGlyLeuTyrAspGly 501
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

us-09-367-496-8.std.rni

Tue Jul 31 13:08:19 2001

```

2278 ATACGAT...GCAACATCACTTTTGTCTCAAGCGGCTTATGACAAA 2324
502 ProValHisGluValMet.....ValProAlaLysPr 512
::: ||| ::::: |||::: |||::: |||::: |||
2325 GGCATTAAAGAAAGAAATAGGACTTGAAAGACAAAGTGTGCGGTAATAA... 2372
512 oGlySerGlyAlaProAlaArgAlaSerCysProGlyLysLeuSerValP 529
::: |||
2373 .....AATGTC..... 2378
529 roProValArgAsnLeuHisGlnSerGlyPheSerLeuSerGlySerGln 545
|||||::: ::::: |||::: |||::: |||
2379 .....AGAAATATCACTTAAACACGCAATCAATTAACACACACTACT 2420
546 AlaAspAspHisLeuAlaArgArgThr.....AlaG1 556
|||::: |||::: |||::: |||::: |||
2421 GCTCACATTGAAGTCAATCTGAACTTACCATGTGTCGTGGATGGCAA 2470
556 nLysIleMetaProProGlyGlyArgSerAsnIleThrSerLeu 571
::: |||::: |||::: |||::: |||::: |||
2471 AGAAGTAACTTCTAAACACGCAATCAATTAAGTGGCGCAACTC 2516

seq_name: /cgn2_6/ptodata/1/1na/PCTUS_COMB.seq: PCT-US96-05800-1

seq_documentation_block:
; Sequence 1, Application PC/TUS9605800
; GENERAL INFORMATION:
; APPLICANT: Oravax, Inc.
; TITLE OF INVENTION: MULTIMERIC, RECOMBINANT UREASE VACCINE
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/05800
; FILING DATE: 23-APR-1996
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/431,041
; FILING DATE: 28-APR-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/568,122
; FILING DATE: 06-DEC-1995
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Clark, Paul T.
; REGISTRATION NUMBER: 30,162
; REFERENCE/DOCKET NUMBER: 06132/020001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-8906
; TELEFAX: (617) 542-8906
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2735 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; PCT-US96-05800-1

alignment_scores:
Quality: 136.50 Length: 632

```

```

Ratio: 0.493 Gaps: 29
Percent Similarity: 43.829 Percent Identity: 19.146

alignment_block:
US-09-367-496-8 x PCT-US96-05800-1 ..
Align seg 1/1 to: PCT-US96-05800-1 from: 1 to: 2735

17 LeuLeuIleArgGlyGlyArgIleValAsnAspGlnSerPheThrAl 33
|||||::: |||::: |||::: |||::: |||
1023 TTAATTATCACTAACGCTTTAATCTGGTGGTACACCGGTATTATTAAGC 1072
::: |||::: |||::: |||::: |||::: |||
33 aspValHisValGluAspGlyLeuIleLysGlnIleGlyGlu..... 47
|||||::: |||::: |||::: |||::: |||
1073 GGATATTGGTTATTAAAGATGCGCAAAATCGCTGGCATTTGTTAAAGCGGTA 1122
::: |||::: |||::: |||::: |||::: |||
48 .....AsnLeuIleValProGlyGly 54
|||||::: |||::: |||::: |||::: |||
1123 ACAAGACATGCAAGATGGGTTTAAACAACTCTAGCGTAGGTCCTGCT 1172
::: |||::: |||::: |||::: |||::: |||
55 IleHisThrIleAspAlaHisGlyLeuMetValLeuProGlyGlyValAs 71
|||||::: |||::: |||::: |||::: |||
1173 ACTGAAGCCTTAGCCGGTGAAGGTTTGCATGCTAACGGCTGGTGTATTGA 1222
::: |||::: |||::: |||::: |||::: |||
71 pValHisThrArgLeuGlnMetProValLeuGlyMetThrProAlaAspA 88
|||||::: |||::: |||::: |||::: |||
1223 CACACACATCCACTTCATTTCCACC..... 1247
::: |||::: |||::: |||::: |||::: |||
88 spPheCysGlnGlyThrLysAlaAlaLeuAlaGlyGly...ThrThrMet 103
|||||::: |||::: |||::: |||::: |||
1248 .....CAACAAATCCCTACAGCTTTTGCAGCGGTGTAAACAACCATG 1289
::: |||::: |||::: |||::: |||::: |||
104 IleLeuAspHisValPheProAspThrGlyValSerLeuLeuAlaAlaTy 120
|||||::: |||::: |||::: |||::: |||
1290 ATTTGGTGGTGAACCGGTCCTGCTGATGCACATAATGCG..... 1328
::: |||::: |||::: |||::: |||::: |||
120 rGluGlnTrpArgGluArgAlaAspSerAlaAlaCysCysAspTyrSerL 137
|||||::: |||::: |||::: |||::: |||
1328 ..... 1328
137 euHisValAspIleThrArgTrpHisGluSerIleLysGluGluLeuGlu 153
::: |||::: |||::: |||::: |||::: |||
1329 .....ACTACTATCACTCCAGGCGAAGAAATTTAAATGGATGCTCAGA 1373
::: |||::: |||::: |||::: |||::: |||
154 AlaLeuValLysGluLysGlyValAsnSerPheLeuValPheMetAlaTy 170
|||||::: |||::: |||::: |||::: |||
1374 GCG...GCTGAAGAATATTCTATGAAT..... 1397
::: |||::: |||::: |||::: |||::: |||
170 rLysAspArgCysGlnCysSerAspSerGlnMetTyrGluIlePheSerI 187
|||||::: |||::: |||::: |||::: |||
1397 ..... 1397
187 leIleArgAspLeuGlyAlaLeuAlaGlnValHisAlaGluAsnGlyAsp 203
::: |||::: |||::: |||::: |||::: |||
1398 .....TTAGGTTTCTGGCTAAAGTAAAGTAAAGTAAAGTAAAGTAAAG 1436
::: |||::: |||::: |||::: |||::: |||
204 IleValGluGluGluGlnLysArgLeuLeuGluLeuGlyIleThrGly... 219
::: |||::: |||::: |||::: |||::: |||
1437 ACCTTAGCCGATCAA.....ATTGAAGCGGTCGATGGCTT 1474
::: |||::: |||::: |||::: |||::: |||
220 .....ProGlu.....GlyHisValL 225
|||||::: |||::: |||::: |||::: |||
1475 TAAATTCACGAAGACTGGGCGACCTCTCTGCAATCAATCATCGCT 1524
::: |||::: |||::: |||::: |||::: |||
225 euSerHisProGluGluValGluAlaGluAlaValTyrArgAlaValThr 241
|||||::: |||::: |||::: |||::: |||
1525 TAGATGTTCCGGACAAATACGATGTGCAAGTGCCTATCCACACACACT 1574
::: |||::: |||::: |||::: |||::: |||
242 IleAlaLysGlnAlaAsnCysProLeuTyrValThrLysValMetSerLys 258
::: |||::: |||::: |||::: |||::: |||
1575 TTG...AATGAAGCGGTTGT..... 1592
::: |||::: |||::: |||::: |||::: |||
258 sGlyAlaAlaAspAlaIleAlaGlnAlaLysArgArgGlyValValValP 275

```

```

1593   ... GTAGACACTATGCTCTATTGCTGACGCACTATGACACT 1638
275   heGlyGluProIleThrAlaSerLeuGlyThrAspGlySerHisThrTp 291
1639 TC .....CACACTGAAGGCGTGGCGGGGACAC ..... 1667
292 SerLysAsnTrpAlaLysAlaAlaPheValThrSerProProValAs 308
1668 GCTCTGTATATTATTAAAGTACCGGTGAACACAACTTCTCCCGCTTC 1717
308 nProAspPro .....ThrThrAlaAspHisLeu 318
1718 CACTAACCCCACTCCCTTACCGTGAATACAGACGACGACATGG 1767
318 hrCysLeuLeuSerSerGlyAspLeuGlnValThrGlySerAlaHisCys 334
1768 ACATGCTTATGCTGTGCCACCACTTGGATAAAGCACTTAAAGAGATGTT 1817
335 ThrPheThrThrAla .....GlnLysAlaValGlyLysAspAs 347
1818 CAGTTCGCTGATTCAAGATCCGCGCTCAACCACTTGCCTGAAGACAC 1867
347 n .....PheAlaLeuIleProGluGlyThrAsnGly 357
1868 TTTGCATGACATGGGATTTTCAATCACCACTTCTGACTCTCAAGCGA 1917
358 .....IleGluGluArgMetSerMetValTrpGluLysCysValAla 371
1918 TGGCCGCTGTGGTGAAGTTTATCACTAGAACTTGGCAACACGCTGACAA 1967
372 Ser .....GlyLysMetAspGluAsnGlu ..... 379
1968 AACAGAAAGAAATTTGCGCGCTGAAGAGAAAGAAAGCGGATGAACGACAA 2017
380 .....PheValAlaValThrSerThrAsnAlaAlaLysI 391
2018 CTTCAAGTCAACAGCTACTTCTCTAAATACACCACTTAACCCAGGATCG 2067
391 lePheAsnPheTrpProArgLysGlyArgValAlaValGlySerAspAla 407
2068 CTCATGGGATTAACGAGTATGAGTTCACTAGAGAGTGGGCAAGTGGCT 2117
408 AspLeuValIleTrpAsnProLysAlaThrLysIleLeuSerAlaLysTh 424
2118 GACTTGGTATGTGGAGTCCA ..... 2138
424 rHisAsnLeuAsnValGluTrpAsnIlePheGluGlyValGluCysArg 441
2139 .....GCATCTTTGGCGTGA ..... 2156
441 lyAlaProAlaValIleSerGlnGlyArgValAlaLeuGluAspGly 457
2157 .....CCCAACATGATCATCAAGCGGATTCATGCTGTTAAGCCAA ..... 2198
458 LysMetPheValThrProGlyAlaGlyArgPheValProArgLysThrPh 474
2199 .....ATGGCGATGCGACGCTTATCCCTACC ..... 2228
474 eProAspPheValTrpLysArgle .....LysAlaAla 485
2229 .CCACACCGGTTTATACAGAGAAATGTTGCTCATCATGCTAAGACTA 2277
485 rgAsnArgLeuAlaGluIleHisGlyValProArgGlyLeuTrpAspGly 501
2278 AATACGAT...GCAACATCACTTTTGTCTCTCAAGCGGCTTATGACAAA 2324
502 ProValHisGluValMet .....ValProAlaLysPr 512
2325 GGCATTAAAGAAATAGGACTTGAAGACAAAGTGTTCGCGGTAA ..... 2372
512 oglySerGlyAlaProAlaArgAlaSerCysProGlyLysIleSerValP 529

```

```

2373 .....AATTGC ..... 2378
529 roProValArgAspLeuHisGlnSerGlyPheSerLeuSerGlySerGln 545
2379 .....AGAAATATCACTAAATAAAGACATGCAATTCAACGACACTACT 2420
546 AlaAspAspHisIleAlaArgArgThr .....AlaG1 556
2421 GCTCACATTGAAGTCAATCTCTGAACCTTACCATGTCTTCTGGATGGCAA 2470
556 nLysIleMetAlaProProGlyGlyArgSerAsnIleThrSerLeu 571
2471 AGAAGTAACTCTTAACACCCCAATAAAGTGAAGCTTGGGCGCAACTC 2516
seq_name: /cgn2_6/ptodata/1/ina/6A_COMB.seq:US-09-036-987A-1

seq_documentation_block:
; Sequence 1, Application US/09036987A
; Patent No. 6143526
; GENERAL INFORMATION:
; APPLICANT: Baltz, Richard H.
; APPLICANT: Broughton, Mary C.
; APPLICANT: Crawford, Kathryn P.
; APPLICANT: Madduri, Krishnamurthy
; APPLICANT: Merlo, Donald J.
; APPLICANT: Treadway, Patti J.
; APPLICANT: Turner, Jan R.
; APPLICANT: Waldron, Clive
; TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
; TITLE OF INVENTION: Production
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dow AgroSciences LLC Patent Department
; STREET: 9330 Zionsville Road
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: USA
; ZIP: 46268
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: 09-MAR-1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Stuart, Donald R
; REGISTRATION NUMBER: 28,479
; REFERENCE/DOCKET NUMBER: 50,608
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (317)337-4816
; TELEFAX: (317)337-4847
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 80161 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-09-036-987A-1

```

alignment_scores:

```

Quality: 122.50 Length: 529
Ratio: 0.513 Gaps: 27
Percent Similarity: 45.180 Percent Identity: 21.739

```

alignment_block:

```

US-09-367-496-8 x US-09-036-987A-1 ..

```

```

Align seg 1/1 to: US-09-036-987A-1 from: 1 to: 80161

```

345 LysAspAsnPheAlaLeuIleProGluGlyThrAsnGlyIleGluGluAr 361
36853CCCTGGGGCTTCCGGCGCGCTGTCGACGGCGCTCGCGGCCCA 36896
361 gMetSerMetValTrpGluLysCysValAlaSer.....GlyLysMetA 376
36897 GGCCGACGGCTCGCGAGCTGTCCGAGCAGACCGATCCCGACCCCGTCG 36946
376 spGluAsnGluPheValAlaValThrSerThrAsnAlaAlaLysIlePhe 392
36947 ATGTCCGGTTCTTCTACTGCCGCCACGGCGACG..... 36978
393 AsnPheTyrProArgLysGlyArgValAlaValGlySerAspAlaAspLe 409
36978 36978
409 uValIleTrpAsnProLysAlaThrLysIleSerAlaLysThrHisA 426
36979GCTTGGGAGCACCGACGGTG...GTCTTGGTCGGGACGAGCGCTA 37021
426 snLeuAsnValGluTyrAsnIlePheGluGlyValGluCysArgGlyAla 442
37022 CGTTCGGCTCCGGCTT.....GGCGTTTTCGCCGCGGTGAA 37059
443 ProAlaValIleSerGlnGlyArgValAlaLeuLysLysPheLysMe 459
37060 CCAGCGGTGATGCTGTTGAGGGAGCGTC...CTGGACGGCAGGTCGT 37106
459 tPheValThrProGly.....AlaGlyArgPheVal.Pr 470
37107 CTTCGTTCTCCCGGTCAGGGCTGGCAGTGGCGCGGTATGCGAGTCGACC 37156
470 oArgLysThrPheProAspPheValTyrLysArgIleLysAlaArgAsnA 487
37157 TGCTGGAGCGTTCCCGCAGCGTTCGGCGCCACATGAGCAGTGCGCCACC 37206
487 tGLeuAlaGluIleHisGlyValProArgGlyLeuTyrAspGlyProVal 503
37207 GCGCTGCGGAGGTACGTGACTGCTTGTTCGACGTGCTCGCGGAGC 37256
504 HisGluValMetValProAlaLysProGlySerGlyAlaProAlaArgAl 520
37257 GGAGAA...CTCCCGACCGCTGGACCGGTGGAGTGTCTCCAGCCGCGT 37303
520 aserCysProGlyLysIleSerValProValArgAsnLeuHisGlns 537
37304 C..... 37304
537 erGlyPheSerLeuSerGlySerGlnAlaAspAspHisIleAlaArgArg 553
37305CTTCGCGT.....GATGGTGTGCTCGCTCGCCGA 37331
554 ThrAlaGlnLysIleMetAlaProProGlyGlyArg 565
37332 GGTTGGCGGTTCTACGGGGTGGCGCGCGCGCT 37367
seq_name: /cgn2_6/ptodata/1/ina/backfiles1.seq:5206163-2
seq_documentation_block:
; Patent No. 5206163
; APPLICANT: RENARD, ANDRE; DINA, DINO; MARTIAL, JOSEPH
; TITLE OF INVENTION: DNA ENCODING BOVINE DIARRHEA
; VIRUS PROTEIN
; NUMBER OF SEQUENCES: 3
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/550,816
; FILING DATE: 06-JUL-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 331,037
; FILING DATE: 29-MAR-1989
; APPLICATION NUMBER: 752,981
; FILING DATE: 08-JUL-1985
; SEQ ID NO:2

5206163-2
LENGTH:12492

alignment_scores:

Quality: 114.50 Length: 360
Ratio: 0.748 Gaps: 13
Percent Similarity: 42.500 Percent Identity: 20.556

alignment_block:

US-09-367-496-8 x 5206163-2

Align seg 1/1 to: 5206163-2 from: 1 to: 12492

```

125 GluArgAlaAspSerAlaAlaCysCysAspTyrSerLeuHisValAspIle 141
10678 GAAATACAGCAGCCAGCCAGGATCATGATAAATTTAGATGATCTTTCA 10727
141 eThrArgTrpHisGluSerIleLys
10728 CACAATAGCCCAACCCAGCCTAAGCACACTTACGCGAAGTGAGCTGGG 10777
150 LuGluLeuGluAlaLeuValLysGluGluValAsnSerPheLeuVal 166
10778 AACAGCTTGAGGAGGAGGATCAACAGAAAGGCGCTGCAGGCTTTCTA... 10824
167 PheMetAlaTyrLysAspArgCysGlnCysSerAspSerGlnMetTyrG 183
10825 .....GAAAAGAGAATCTTGAGAGAGTACTGGGACTCAGAGAAGCACCT 10868
183 uilePheSerIleIleArgAspLeuGluAlaLeuAlaGlnValHisAla. 199
10869 GGTGGACCACTAATCAGAGACCTGAAAACAGAGGAGGAGATAGATATT 10918
199 .....
10919 ATGAGACAGCAATACCTAAGAACAGAGAGAGGATGTCAGTGAGATGG 10968
200 GluAsnGlyAspIleValGluGluGlnLysArgLeuLeuGluLeuG 216
10969 CAAGCAGGGGACATAGTTGATGAAGAAAGAACCAAGA..... 11004
216 ylleThrGlyProGluGlyHisValLeuSerHisProGluLeuValGlu 233
11005 .....GTGATCAATACCTGAA..... 11022
233 laGluAlaValTyrArgAlaValThrIleAlaLysGlnAlaAsnCysPro 249
11023 .....GCTAAGACAGA 11034
250 LeuTyrValThrLysValMet..... 256
11035 CTGGCCATCACTAAAGTTATGTACAACTGGGTGAGCAGCAGCCCTGTTGT 11084
257 .....SerLysGlyAlaAlaAspAlaIleAlaGlnAlaL 268
11085 GATCCAGGATGAAGGAGGAGAGCCCATTTATCAAGATCTTTACACAGG 11134
268 ysArgArgGlyValValPheGlyGluProIleThrAlaSerLeuGly 284
11135 TAAGAAAGGAATGGGACCTGTTTCATGAGCCAGTGTGAGTTTTCAT 11184
285 ThrAspGly..... 287
11185 ACTAAGGCTGGACACCAAGTCACTAGTAGGATCTACGGCTTATTGG 11234
288 .....SerHisTyrTrpSerLysAsnTyrAlaLysAlaAlaLap 301
11235 TGAATTCAAAATATATTACTACAGAGGAGTGCACAAATTCATCGATA 11284
301 heValThrSerProValAsnProAspProThrThrAlaAspHisLeu 317
11285 CCATCAGCAGCACATGGTGGGTACCTGATCATACAGCAGAT..... 11328

```

```

318 ThrCysLeuLeuSerSerGlyAspLeuGlnValThrThrGlySerAlaHisCy 334
11329 .....GTTGAAGTATACATA..... 11343
334 sThrPheThrThrAlaGlnLysAlaValGlyLysAspAsnPheAlaLeuI 351
11344 .....AGAAATGGACAAGGGGTAGTGGCCAGCAGAC..... 11376
351 leProGluGlyThrAsnGlyIleGluGluArgMetSerMetValTrpGlu 367
11377 ..ACAAGCGCAGGTAAACAGCATGCTAAATGTGTAAACAATGATGATGCC 11424
368 LysCysValAlaSerGlyLysMetAspGluAsnGluPheValAlaValTh 384
11425 TTCTGTGAAGTAGCGGGGTT..... 11445
384 rSerThrAsnAlaAlaLysIlePheAsnPheTyrProArgLysGlyArgV 401
11446 .....CCATATAAGAGTTTAAAT.....AGAGTTGCAAGGA 11476
401 alaAlaVal...GlySerAspAlaAspLeuValIleTrpAsnProLysAla 416
11477 TCATGCTGTGGGGATGACGCTTCCTGATAACAGAGAGGGGCTGGGC 11526
417 ThrLysIleIleSerAlaLysThrHisAsn 426
11527 ACTAAATTTGCCAACAAAGGGATGCAAC 11556
seq_name: /cgn2_6/ptodata/1/lna/6A_COMB.seq:US-09-320-878-19
seq_documentation_block:
; Sequence 19, Application US/09320878A
; Patent No. 6117659
; GENERAL INFORMATION:
; APPLICANT: ASHLEY, Gary
; APPLICANT: BETLACH, Melanie C.
; APPLICANT: BETLACH, Mary C.
; APPLICANT: MCDANIEL, Robert
; APPLICANT: TANG, Li
; TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 300622002120
; CURRENT APPLICATION NUMBER: US/09/320,878A
; EARLIER FILING DATE: 1999-05-27
; EARLIER APPLICATION NUMBER: CIP OF 09/141,908
; EARLIER FILING DATE: 1998-08-28
; EARLIER APPLICATION NUMBER: CIP OF 09/073,538
; EARLIER FILING DATE: 1998-05-06
; EARLIER APPLICATION NUMBER: CIP OF 08/846,247
; EARLIER FILING DATE: 1997-04-30
; EARLIER APPLICATION NUMBER: 60/119,139
; EARLIER FILING DATE: 1999-02-08
; EARLIER APPLICATION NUMBER: 60/100,880
; EARLIER FILING DATE: 1998-09-22
; EARLIER APPLICATION NUMBER: 60/087,080
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19
; LENGTH: 38506
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-320-878-19

```

alignment_scores:

Quality: 112.50 Length: 478
Ratio: 0.511 Gaps: 26
Percent Similarity: 46.025 Percent Identity: 23.640

alignment_block:

US-09-367-496-8 x US-09-320-878-19

us-09-367-496-8.std.rni

Tue Jul 31 13:08:19 2001

```

Align seg 1/1 to: US-09-320-878-19 from: 1 to: 38506
178 AspSerGlnMetTyrGluLeuPheSerIleIleArgAspLeuGlyAlaLe 194
||||| : : : : : |||||||
17421 GACTCGCGCTCGCGGGCTGTCGCGGAACCTCGCGGACCTGGCGCGAC 17470
||||| : : : : : |||||||
194 uAlaGlnValHisAlaGluAsnGlyAspIleValGluGluGlnLysA 211
||| ||| : : : : : |||
17471 GGCACCGTC.....GTGACCTCGACCTCAGCGAGCGGAGCGCGCG 17514
||| ||| : : : : : |||
211 rglLeuGluLeuGlyIleThrGly.....ProGluGlyHisValLeu 225
||| ||| : : : : : |||
17515 CCCGCTGCTCGCGGGCTTCGACGACGCGACCGCTCAGCGCGCTCTC 17564
||| ||| : : : : : |||
226 SerHisProGluGluVal.....GluAlaGln 234
||| ||| : : : : : |||
17565 CACCTGCGCGCCACCGCTCGACTCGAGCGCTCGCGCGACCGCGGA 17614
||| ||| : : : : : |||
234 uAlaValTyrArgAlaValThrIleAlaLysGlnAlaAsnCysProLeu 251
||||| : : : : : |||
17615 CGCGCTCGCGCTGTCGTGAC.....GCGAGGCGCCACCGCGCGCTCC 17658
||| ||| : : : : : |||
251 yrValThrLysValMetSerLysGlyAlaAlaAspAlaIleAlaGln 267
||| ||| : : : : : |||
17659 ACCTGACCGCGCTCTCG.....CGGAGCGCGCGCTCGCGGA 17696
||| ||| : : : : : |||
268 LysArgArgGlyValValPheGlyGluProIleThrAlaSerLeu 284
||| ||| : : : : : |||
17697 GCGCGTCCGCGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 17746
||| ||| : : : : : |||
284 yThrAspGlySerHisTyrTyrSerLysAsnTyrAlaLysAlaAla 301
||| ||| : : : : : |||
17747 CGCGCGCGGTACGGCGCTACGCGCGGTACGCGCTCTCTCTCTCTCT 17796
||| ||| : : : : : |||
301 he.....ValThrSerProVal 307
||| ||| : : : : : |||
17797 TCGCGGTACGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 17846
||| ||| : : : : : |||
308 AsnPro..... 309
17847 AGCCCTTGGAGGCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCT 17896
||| ||| : : : : : |||
310 .....AspProThrThrAlaAspHisLeu 318
||| ||| : : : : : |||
17897 CGCGCGCTCGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 17940
||| ||| : : : : : |||
318 hrCysLeu.....LeuSerSerGlyAspLeuGlnValThr..... 329
||| ||| : : : : : |||
17941 CCGCGCTGGACACCGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCG 17990
||| ||| : : : : : |||
330 .....GlySerAlaHisCysThrPheThrThrAlaGlnLysAl 342
||| ||| : : : : : |||
17991 GACGTGCGACTGGTGGAGCTTCGCGCGCGCGCGCGCGCGCGCGCG 18040
||| ||| : : : : : |||
342 aValGlyLysAspAsnPheAlaLeuIleProGluGlyThrAsnGlyIle 359
||| ||| : : : : : |||
18041 CACC.....CTCTCGCGATGTCGCGCGCGCGCGCGCGCGCGCG 18081
||| ||| : : : : : |||
359 luLuArgMetSerMetValTyrGluLysCysValAlaSerGlyLysMet 375
||| ||| : : : : : |||
18082 ACAGCAGCAGTCGACGCG.....GCGCGCGCGCGCGCGCGCGCTC 18119
||| ||| : : : : : |||
376 AspGluAsnGluPheValAlaValThrSerThrAsnAlaAlaLysIle 392
||| ||| : : : : : |||
18120 CTGAGCGCGCGCGCTCGGTGCTCAGCGCGCGCGCGCGCGCGCG 18155
||| ||| : : : : : |||
392 eAsnPheTyrProAlaGlyGlyArgValAlaValGlySerAspAlaAsp 409
||| ||| : : : : : |||
18156 .....CAGCAGCGCGGTATCAGGAGGTGTCGCGCGCGCGCGCG 18198
||| ||| : : : : : |||
409 euValIleTrpAsnProLysAlaThrLysIleSerAlaLysThrHis 425
||| ||| : : : : : |||
18199 TGGTCTCAACCAACCGCTCCCGCGAGGCGGTGTCGACACGCGGCG 18248
||| ||| : : : : : |||

```

seq_name: /cgn2_6/ptodata/1/ina/6A_COMB.seq:US-09-000-016-1

seq_documentation_block:

; Sequence 1, Application US/09000016

; Patent No. 6143541

; GENERAL INFORMATION:

; APPLICANT: Akira ARISAWA et al.

; TITLE OF INVENTION: GENE ENCODING A PROTEIN HAVING ASYMMETRIC

; TITLE OF INVENTION: HYDROLASE ACTIVITY FOR 4-SUBSTITUTED 1,4-DIHYDROPYRIDINE DE

; TITLE OF INVENTION: ITS EXPRESSION PRODUCT

; NUMBER OF SEQUENCES: 7

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.

; STREET: 2033 K Street, N.W., #800

; CITY: Washington

; STATE: D.C.

; COUNTRY: U.S.A.

; ZIP: 20006

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: MS-DOS

; SOFTWARE: Wordperfect 5.1

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/000.016

; FILING DATE: January 30, 1998

; CLASSIFICATION: 530

; PRIOR APPLICATION DATA:

```

alignment_scores:
  Quality: 110.50      Length: 643
  Ratio: 0.423        Gaps: 29
  Percent Similarity: 40.591  Percent Identity: 20.684

alignment_block:
US-09-367-496-8 x US-09-000-016-1/rev ..

Align seg 1/1 to reverse of: US-09-000-016-1 from: 1 to: 2809

7  LysSerIleProArgIleThrSerAspArgLeu...LeuIleArgGlyG1 22
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1633 CGGTCCCTCAGCGACCTCTCGGCCGATGCTGTGCCTCCGCCGAGGCGG 1584
      22 yArgIleValAsnAspAspGlnSerPheTyAlaAspValHisValGlu 39
      ||||| : : : : : : : : : : : : : : : : : : : : : :
1583 CCGT.....: : : : : : : : : : : : : : : : : : : : : :
      39 spGlyLeuIleLys : : : : : : : : : : : : : : : : : :
1563 GCGGTGACGTCCGGCTTGATGCGCGCTGATCGCGGGGCGCGGTGGA 1514
      44 ...GlnIleGlyGluAsnLeuIleValProGlyGlyIleHisThrIleAs 59
      : : : : : : : : : : : : : : : : : : : : : : : : : :
1513 GGAGAAGTCGCGGAGCTGCTCTGCTGCACGGC.....GCCGA 1473
      59 pAlaHisGlyLeuMetValLeuProGlyGlyValAspValHisThrArgL 76
      ||||| : : : : : : : : : : : : : : : : : : : : : :
1472 CGGTGAGGGGGCGCTCGCGCTGCGGGCGAACCGAT.....CGAC 1432
      76 euGlnMetProValLeuGlyMetThrProAlaAspAspPheCysGlnGly 92
      ||||| : : : : : : : : : : : : : : : : : : : : : :
1431 TCGGGGCGCTCTGTTCGGCGCGC.....GATGCC 1403
      93 ThrLysAlaLeuAlaGlyGlyThrThrMetIleLeuAspHisValPh 109
      : : : : : : : : : : : : : : : : : : : : : : : : : :
1402 GAACAGGACGCCCTCTTCGGCGGA.....CAGCTTGTGACCCCGCGCT 1359
      109 eProAspThrGlyValSerLeuLeuAlaAlaTyArgIleGlnTrpArgGlu 126

```

* Tue Jul 31 13:08:19 2001

```
698 CGAGTTCGGTGATGTCGAGAGCGCGCGGTCGAGCTTGCCG..... 658
;
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: USA
; ZIP: 46268
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/036,987A
; FILING DATE: 09-MAR-1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Stuart, Donald R
; REGISTRATION NUMBER: 28,479
; REFERENCE/DOCKET NUMBER: 50,608
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (317)337-4816
; TELEFAX: (317)337-4847
;
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 80161 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-09-036-987A-1

alignment_scores:
  Quality: 110.00      Length: 685
  Ratio: 0.364        Gaps: 35
  Percent Similarity: 44.088      Percent Identity: 20.146

alignment_block:
US-09-367-496-8 x US-09-036-987A-1/rev ..

Align seg 1/1 to reverse of: US-09-036-987A-1 from: 1 to: 80161

21 GlyGlyArgIleValAsnAspGlnSerPheTyrAlaAspValHisVa 37
   ||||| : : : : : ||||| : : : : : |||||
9704 GCGGGAATGCTGGACGGATGCAACAGCAGCACCGTCCGGGTGTCATGT 9655
   ||||| : : : : : ||||| : : : : : |||||
37 lGluAspGlyLeuIleLysGlnIleGlyGluAsnLeuIleValProGly 54
   ||||| : : : : : ||||| : : : : : |||||
9654 TGACGATGTTGCTCTGAC.....GAAGGAGAGAGCAGCACCTGGCC 9615
   ||||| : : : : : ||||| : : : : : |||||
54 lylleHisThrIleAspAlaHisGlyLeuMetValLeuProGlyGlyVal 70
   ||||| : : : : : ||||| : : : : : |||||
9614 GAGCGTCAACCCAGAGAAAGTCTGCGAGGACTGCGACTTCTCTCGGACTT 9565
   ||||| : : : : : ||||| : : : : : |||||
71 .....AspValH1 73
   ||||| : : : : : ||||| : : : : : |||||
9564 CCACCACCATGTTCCGGTTCGGTTCGGTAGAACCCAGGACCCCTGTCTCA 9515
   ||||| : : : : : ||||| : : : : : |||||
73 sThrArgLeuGlnMetProValLeuGlyMetThrProAlaAspPheC 90
   ||||| : : : : : ||||| : : : : : |||||
9514 GACTGGAGCAGCTTACACGACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 9475
   ||||| : : : : : ||||| : : : : : |||||
90 ysGlnGlyThrLysAlaAlaLeuAla.GlyGlyThrThrMetIleLeuAs 106
   ||||| : : : : : ||||| : : : : : |||||
9474 ..CGAGGAGTAGTCCACATAGGCGGGAACGCTGCCACGGTGTGCGCTCG 9427
   ||||| : : : : : ||||| : : : : : |||||
106 phiValPheProAspThrGlyValSerLeuLeuAlaAlaTyrGluGlnT 123
   ||||| : : : : : ||||| : : : : : |||||
9426 TGAGTGTCTCGG.....AGTTGCTCTGACCCGTC 9398
   ||||| : : : : : ||||| : : : : : |||||
123 rfaGluArgAlaAspSerAlaAlaCysCysAspTyrSerLeuHisVal 139
   ||||| : : : : : ||||| : : : : : |||||
9397 GCCGAGAGCTGCAGGAC.....GTT 9378
   ||||| : : : : : ||||| : : : : : |||||

;
; STREET: 9330 Zionsville Road
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: USA
; ZIP: 46268
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/036,987A
; FILING DATE: 09-MAR-1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Stuart, Donald R
; REGISTRATION NUMBER: 28,479
; REFERENCE/DOCKET NUMBER: 50,608
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (317)337-4816
; TELEFAX: (317)337-4847
;
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 80161 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-09-036-987A-1

seq_name: /cgn2_6/ptodata/1/ina/6A_COMB.seq:US-09-036-987A-1

seq_documentation_block:
; Sequence 1, Application US/09036987A
; Patent No. 6143526
; GENERAL INFORMATION:
; APPLICANT: Baltz, Richard H.
; APPLICANT: Broughton, Mary C.
; APPLICANT: Crawford, Kathryn P.
; APPLICANT: Madduri, Krishnamurthy
; APPLICANT: Merlo, Donald J.
; APPLICANT: Treadway, Patti J.
; APPLICANT: Turner, Jan R.
; APPLICANT: Waldron, Clive
; TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
; NUMBER OF INVENTION: Production
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dow Agrosciences LLC Patent Department
```

[illegible]

```

8634 ..... |||||:.. CCTCGTGGAAATCGGTTGGAGC 8612
403 1.....
8611 GCAACCTCTGGGGGAGATCAACACCCCATTTGGAGATCTCTGGAAGAAT 8562
407 laAspLeuValIleTAsnProLysAlaThrLysIleIleSerAlaLys 423
8561 CTGACTTC.....GGACTCGCTGGCATCGTCCCGG 8530
424 ThrHisAsnLeuAsnValGluTyrAsnIlePheGluGlyValIcuCysAr 440
8529 TCACGGCGGGAATCCGAGTCGACGACACGCGATGAACACTGCCA..... 8485
440 gGlyAlaProAlaValIleSerGlnGlyArgValAlaLeuGluAspG 457
8484 ...AACCCCGCGCTCCATGTTTC...GGCCACACGTTGCTACGGATG 8442
457 ly.....LysMetPheValThr 462
8441 GCATCAGACTGCCCAGTCGCGCGGTTCTCAGTTCGCAACTACGCTTTT 8392
463 ProGlyAlaGlyArg.PheValProArgLysThrPheProaspPheVal. 478
8391 CCAGACGTGGACGATGTCGTGGCGGTGGGACTTCTCGATTTCACGG 8342
479 .....TyrLysArgIleLysAlaArgAsnArgLeuAlaGlu..... 490
8341 CGGATCAGCGCGGTTGGCAGCAGCAACGCGCTCGTCGACGCGCGGC 8292
491 .....IleHisGlyValProArgGlyLeuTyrAsp.. 500
8291 CAGCTGACCTCGCAGCTCGTCTTCGGTAAACGTCGCGGGGTGCGAGSACCA 8242
501 .....GlyProValHisGluValMetValProAlaLys.ProGlySer 514
8241 GACCGGTCCCGATCGGTCAGAGCTCCGCCCTACGAGATTCGTCGCCAG 8192
515 GlyAlaProAlaArgAlaSerCysProGlyLysIleSerValProPro.. 530
8191 AAGTCCCGAGGAGATCACTCGGTGTACCGCTGTACCGTGGCGGTTTC 8142
531 .....ValArgAsnLeuHisGlnSerGlyP 539
8141 CTGCGTCGTCTGACGCCATGATGGATGATCACTGAACACGACTCCAGCA 8092
539 heSerLeuSerGly..... 543
8091 GTTCGTTGACGGTACGTATTCTGTGACCGCGAGCTTCGAGGCGCACTCC 8042
544 .....SerGlnAlaAspAspHisIleAlaArgArgThrAlaGln 556
8041 CCATCTCCCGTACTTCGTCGACAGCAAGGTGGCGATCACCTCCGACGTC 7992
556 nLysIleMetAlaProProGlyClyArgSerAsnIleThrSerLeuSer 572
7991 GAGCTGGCGCGGCCG.....CGCAACCACTTTCCACCATGTGT 7952

```

us-09-367-496-8.std.rni

Tue Jul 31 13:08:19 2001

Tue Jul 31 13:08:18 2001

OM of: US-09-367-496-8 to: N_Geneseq_0601.* out_format : pfs

Date: Jul 30, 2001 12:42 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=framet_p2n.model -DEV=xlpl
-Q=/cgn2_1/USPTO_spool/US09367496/runat_30072001_081221_27409/app_query.fasta_1.637
-DB=N_Geneseq_0601 -OFMT=fastq -SURFIX=std.rng -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-XGAPOP=6.000 -XGAPEXT=7.000 -XGAPOP=10.000 -XGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blomsm62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -MINLEN=0 -MAXLEN=2000000000
-USER=US09367496_CGNI_1_224 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
-WAIT -THREADS=1

Search information block:

Query: US-09-367-496-8
Query length: 572
Database: N_Geneseq_0601.*
Database sequences: 730101
Database length: 313950809
Search time (sec): 112.950000

score_list:	Strd Orig	zScore	EScore	Len	Documentation
/SID88/gcgdata/geneseq/geneseq/NA1998.DAT:AAV60818		2879.00	4602.51	1690	2.2e-248
/SID88/gcgdata/geneseq/geneseq/NA1998.DAT:AAV60817		2803.00	4478.99	1920	1.6e-241
/SID88/gcgdata/geneseq/geneseq/NA1998.DAT:AAV60815		2325.00	3711.45	1817	9.3e-199
/SID88/gcgdata/geneseq/geneseq/NA1998.DAT:AAV60816		2151.00	3429.29	2297	4.8e-183
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV60842		1044.00	1651.59	2052	5.0e-84
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV60879		1018.00	1651.59	1734	5.8e-82
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV60092		1011.50	1602.29	1569	2.8e-81
/SID88/gcgdata/geneseq/geneseq/NA1995.DAT:AAV61258		837.00	1319.57	1938	1.6e-65
/SID88/gcgdata/geneseq/geneseq/NA1995.DAT:AAV60819		826.00	1305.57	1383	9.5e-65
/SID88/gcgdata/geneseq/geneseq/NA1987.DAT:AAV608565		805.00	1268.46	11e-62	13883
/SID88/gcgdata/geneseq/geneseq/NA1995.DAT:AAV608154		698.50	1076.22	5.6e-52	13085
/SID88/gcgdata/geneseq/geneseq/NA1995.DAT:AAV613316		691.50	1089.50	1.0e-52	1374
/SID88/gcgdata/geneseq/geneseq/NA1995.DAT:AAV613256		684.50	1071.66	1.0e-51	2518
/SID88/gcgdata/geneseq/geneseq/NA1996.DAT:AAV613257		684.50	1069.09	1.4e-51	3190
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV621851		600.00	933.82	4.8e-44	3039
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV606146		415.00	662.96	5.9e-29	268
/SID88/gcgdata/geneseq/geneseq/NA1993.DAT:AAV6094181		409.50	636.31	1.8e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094184		409.50	636.31	1.8e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094179		408.50	634.70	2.2e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094168		408.50	645.86	5.3e-28	459
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094178		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094180		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094182		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094183		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094184		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094185		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094186		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094187		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094188		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094189		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094190		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094191		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094192		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094193		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094194		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094195		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094196		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094197		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094198		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094199		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094200		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094201		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094202		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094203		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094204		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094205		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094206		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094207		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094208		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094209		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094210		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094211		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094212		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094213		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094214		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094215		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094216		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094217		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094218		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094219		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094220		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094221		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094222		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094223		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094224		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094225		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094226		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094227		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094228		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094229		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094230		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094231		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094232		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094233		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094234		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094235		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094236		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094237		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094238		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094239		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094240		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094241		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094242		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094243		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094244		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094245		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094246		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094247		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094248		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094249		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094250		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094251		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094252		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094253		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094254		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094255		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094256		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094257		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094258		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094259		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094260		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094261		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094262		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094263		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094264		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094265		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094266		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094267		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094268		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094269		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094270		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094271		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094272		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094273		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094274		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094275		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094276		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094277		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094278		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094279		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094280		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094281		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094282		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094283		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094284		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094285		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094286		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094287		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094288		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094289		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094290		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094291		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094292		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094293		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094294		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094295		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094296		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094297		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094298		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094299		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094300		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094301		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6094302		407.50	633.10	2.7e-27	1377
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:					

1 MetSerPheGlnGlyLysLysSerIleProArgIleThrSerAspArgLe 17
|||||
32 ATGTCTCTCCAGGGCAAGAAAGCATCCCGGATCAGAGTGACGGCT 81
|||||
17 UleuIleArgGlyGlyArgIleValAsnAspAspGlnSerPheTyrAlaA 34
|||||
82 TCTGATCAGAGTGGAGAGTCGTGAATGAGCAGCAGTCTTTACGGTG 131
|||||
34 spValHisValGluAspGlyLeuIleLysGlnIleGlyGluAsnLeuIle 50
|||||
132 ATGTGCAGCTGGAAGATGCTTGTATAAACAATCGGAAACCTCATC 181
|||||
51 ValProGlyGlyIleHisThrIleAspAlaHisGlyLeuMetValLeuPr 67
|||||
182 GTCCCTGGGCACTAGACCATTTGAGCGCCACGGCTGATGGTCTTC 231
|||||
67 oGlyGlyValAspValHisThrArgLeuGlnMetProValLeuGlyMet 84
|||||
232 TGGTGGGGTTGACCTCCACAAAGGCTGCAGATCCCTGCTCTGGGATGA 281
|||||
84 hrProAlaAspAspPheCysGlnGlyThrIysAlaAlaLeuAlaGlyGly 100
|||||
282 CACGGCTGACGACTCTGTACGGGCACCAAGCAGCGCTAGCAGGAGA 331
|||||
101 ThrThrMetIleLeuAspHisValPheProAspThrGlyValSerLeuLe 117
|||||
332 ACCACCATGATCTTGACCAACGCTTCCCGACACGGGTGTGAGCTGCT 381
|||||
117 uAlaAlaTyrGluGlnTyrArgGluArgAlaAspSerAlaAlaCysAla 134
|||||
382 GCGGGCTACGAGCAGTGGCGGAGCGGGGACAGCGCGCTGCTCGC 431
|||||
134 spTyrSerLeuHisValAspIleThrArgTyrPheHisGluSerIleLysGlu 150
|||||
432 ACTACTCCCTGCACGTGGACATCACCCGATGCGATGAGAGCATCAAGAG 481
|||||
151 GluLeuGluAlaLeuValLysLysGlyValAsnSerPheLeuValPh 167
|||||
482 GAGCTGAGGGCTGCTCAAGGAGAAGGGTGTGAACCTCTCTCTGCTT 531
|||||
167 eMetAlaTyrLysAspArgCysGlnCysSerAspSerGlnMetTyrGluI 184
|||||
532 CATGGCATACAAAGACCGGTGCCAGTGCAGCAGCAGCAGCATGTACGAGA 581
|||||
184 lePheSerIleIleArgAspLeuGlyAlaLeuAlaGlnValHisAlaGlu 200
|||||
582 TCTTCAGCATCATCCGGACCTGGGGGCTTGGCCCGAGGTGCACGTGAG 631
|||||
201 AsnGlyAspIleValGluGluGlnLysArgLeuLeuGluLeuGlyIle 217
|||||
632 AACGGGACATCTGGAGGAGGAGCAGAACGGCTTCTCGAGCTCGGCAT 681
|||||
217 eThrGlyProGluGlyHisValLeuSerHisProGluGluValGluAlaG 234
|||||
682 CACTGGCCCGAGGGCCAGTGTCTAGCCACCCCGAGAGGTGGAGGTG 731
|||||
234 luAlaValTyrArgAlaValThrIleAlaLysGlnAlaAsnCysProLeu 250
|||||
732 AGCGGTGTACCGAGTGTCAACATCGCCACAGCAGCAACTGCCCGCTG 781
|||||
251 TyrValThrLysValMetSerLysGlyAlaAlaAspAlaIleAlaGlnAl 267
|||||
782 TAGCTACCAAGGTGATGAGCAAGGGGCGCGCAGCCATCGCTCAGGC 831
|||||
267 aLysArgGlyValValValPheGlyGluProIleThrAlaSerLeuG 284
|||||
832 CAAGCGCAGAGGGTGGTCTGTTTGGGAGCCCATCACCCCGAGCCTGG 881
|||||
284 lyThrAspGlySerHisTyrTrpSerLysAsnTrpAlaLysAlaAlaAla 300
|||||
882 GCACCGAGGTTACACTTACTTGGAGCAAGAACTGGGCCAAGGCTGCAGCC 931
|||||
301 PheValThrSerProProValAsnProAspProThrThrAlaAspHisLe 317

|||||
932 TTCGTACATCACCCCTGTCAACCCAGACCCACCACGCGCACCACT 981
|||||
317 uThrCysLeuLeuSerSerGlyAspLeuGlnValThrGlySerAlaHisC 334
|||||
982 CACTCTGCTGCTCTCCAGCGGGACCTCTCAGGTGACAGGCAGCGCACT 1031
|||||
334 yStrPheThrThrAlaGlnLysAlaValGlyLysAspAsnPheAlaLeu 350
|||||
1032 GCACCTTCACCACTGCCAGAGGCTGTGGCAAGGACAACCTTCGCGCTG 1081
|||||
351 ileProGluGlyThrAsnGlyIleGluArgMetSerMetValTrpGl 367
|||||
1082 ATCCCGGAGGGCACCAACGGCATTTGAGGCGCATGTGATGGTGGGA 1131
|||||
367 uLysCysValAlaSerGlyLysMetAspGluAsnGluPheValAlaValT 384
|||||
1132 GAAATGTGTGGCTCTGGGAAGATGGAGAGAAATGAGTTCGTCGGGTGA 1181
|||||
384 hrSerThrAsnAlaAlaLysIlePheAsnPheTyrProArgLysGlyArg 400
|||||
1182 CCAGTACAAATGCTGCCAAATCTTCAATTTTACCCCAAGGAGGGCGCA 1231
|||||
401 ValAlaValGlySerAspAlaAspLeuValIleTrpAsnProLysAlaTh 417
|||||
1232 GTGCTGTGGGCTCTGACGTGACCTGCTCATATGGAACCCCAAGGCCAC 1281
|||||
417 rLysIleIleSerAlaLysThrHisAsnLeuAsnValGluTyrAsnIleP 434
|||||
1282 CAAGATCATCTTGCCAAAGACCCACAATCTGAACGTGGAGTACAACATCT 1331
|||||
434 heGluGlyValGluCysArgGlyAlaProAlaValIleSerGlnGly 450
|||||
1332 TCAGGAGGTGGAGTGGCGGGAGCGCTGCCGTGCTCATAGTCAGGGC 1381
|||||
451 ArgValAlaLeuGluAspGlyLysMetPheValThrProGlyAlaGlyAr 467
|||||
1382 CGAGTGGCGCTGGAGGACGGGAAGATGTTGTACCCCGGGGGCGCG 1431
|||||
467 gPheValProArgLysThrPheProAspPheValTyrLysArgIleLysA 484
|||||
1432 CTTCGTCCCTCGAAACATTCCTCGACTTGTCTACAAGAGGATCAAG 1481
|||||
484 laArgAsnArgLeuAlaGluIleHisGlyValProArgGlyLeuTyrAsp 500
|||||
1482 CTCGCAACAGGCTGGCGGAGATCCACGGTGTGCCCTGTGTATGAC 1531
|||||
501 GlyProValHisGluValMetValProAlaLysProGlySerGlyAlaPr 517
|||||
1532 GGGCCGTCTCCAGAGGTGATGTGCTGCCAAGCCAGGAGTGGCGCTCC 1581
|||||
517 oAlaArgAlaSerCysProGlyLysIleSerValProProValArgAsnL 534
|||||
1582 GGGCCGCGCTCTCCAGGCAAGATCTCGTCCCTCTCTGTGGCGCAAC 1631
|||||
534 euHisGlnSerGlyPheSerLeuSerGlySerGlnAlaAspAspHisIle 550
|||||
1632 TACATAGTGGGGTTTCCAGCTATCTGGGTCTCAGGCTGATGACCATC 1681
|||||
551 AlaArgArg 553
|||||
1682 GCCGAGCGC 1690

seq_name: /SID8/gcgdata/geneseq/geneseq/NA1998.DAT:AAV60817

seq_documentation_block:

ID AAV60817 standard; cDNA; 1920 BP.

XX AC AAV60817;

XX DT 08-DEC-1998 (first entry)

XX DE Mouse ULIP-4 coding sequence.

Tue Jul 31 13:08:18 2001

XX Mouse; Unc-33-like phospho-protein; ULIP; rat; neoplasm; tumorigenesis;
 KW neurodegenerative disorder; diagnosis; ss.
 XX Mus musculus.

XX Key Location/Qualifiers
 FH 102..1820
 FT /*tag= a
 FT /product= "ULIP-4"
 FT /note= "Unc-33-like phosphoprotein 4"
 XX

FR2759701-A1.

21-AUG-1998.

19-FEB-1997; 97FR-0001961.

19-FEB-1997; 97FR-0001961.

(INEM) INSERM INST NAT SANTE & RECH MEDICALE.

XX Aguera M, Belin MF, Byk T, Honnorat J, Kolattukudy P;
 PI Quach TT, Sobel A;
 XX WPI; 1998-449610/39.
 DR P-PSDB; AAV68488.

XX Mouse and human ULIP poly:peptide(s) - useful in detection of
 PT para-neoplastic neurological syndromes

XX Claim 3; Fig 11; 90pp; French.

XX This sequence represents the mouse Unc-33-like phospho-protein (ULIP)-4
 CC coding sequence. The sequence was isolated based on similarity to the
 CC rat ULIP sequence. Proteins of the ULIP family or their corresponding
 CC nucleic acids can be used in compositions for treating neurodegenerative
 CC disorders and neoplasms, especially for para-neoplastic neurological
 CC syndromes and/or for the early diagnosis of tumorigenesis.

XX Sequence 1920 BP; 440 A; 524 C; 562 G; 394 T; 0 other;

alignment_scores:
 Quality: 2803.00 Length: 572
 Ratio: 4.988 Gaps: 0
 Percent Similarity: 98.252 Percent Identity: 92.657

alignment_block:

US-09-367-496-8 x AAV60817 ..
 Align seg 1/1 to: AAV60817 from: 1 to: 1920
 1 MetSerPheGlnGlyLysLysSerIleProArgIleThrSerAspArgLeu 17
 102 ATGTCCTTCNAGGCAAGAGAGCATTCCTCCGGATAACGAGCGACCGCCT 151
 17 uLeuIleArgGlyGlyArgIleValAsnAspAspGlnSerPheTyrAla 34
 152 TCTCATCAAGGTGGAGATGTGTAACGATGACCGATCCTTCATGCTG 201
 34 spValHisValGluAspGlyLeuIleLysGlnIleGlyGluAsnLeu 50
 202 ATCTGTATGTGAAGACCGCTGATTAACAAATTTGGAGAAATCTCATC 251
 51 ValProGlyGlyIleHisThrIleAspAlaHisGlyLeuMetValLeu 67
 252 GTCCCTGGGGGATCAAAACCATCATGCTCATGCCCTGATGGTGGCGCC 301
 67 oGlyGlyValAspValHisThrArgLeuGlnMetProValLeuGlyMet 84
 302 TGGGGGAGTTGACGTTACACCCCGCTGCAGATGCTGTGATGGCATGA 351

84 hrProAlaAspPheCysGlnGlyThrLysAlaAlaLeuAlaGlyGly 100
 352 CCCACGCTGATGATTTCTGTCAGGCGACCAAGCGGCTCTAGCAGCGGG 401
 101 ThrThrMetIleLeuAspHisValPheProAspThrGlyValSerLeu 117
 402 ACCACCATGATATTGGACCATGTGTTCTGACGCTGTGTGAGCTGCT 451
 117 uAlaAlaTyrGluGlnTrpArgGluArgAlaAspSerAlaAlaCysCys 134
 452 GGCAGCCTATGAGCAGTGGCGGAGGAGCAGCAGCGCGCTGCTGTG 501
 134 spTyrSerLeuHisValAspIleThrArgTyrHisGluSerIleLysGlu 150
 502 ACTACTCTCTTACATGTGGACATTCCTCGCTGGCAGCAGACCAAGAA 551
 151 GluLeuGluAlaLeuValLysGluLysGlyValAsnSerPheLeuValPh 167
 552 GAGCTGGAGGCCCTAGTCAGGAGCAAAAGGTGTGAACCTCTCTTCTG 601
 167 eMetalatyrLysAspArgCysGlnCysSerAspSerGlnMetTyrGlu 184
 602 CATGSCATACAAAGGAGCAGGTGCCAGTGTACTGACGCTCAGATATATA 651
 184 lePheSerIleIleArgAspLeuGlyAlaLeuAlaGlnValHisAlaGlu 200
 652 TCTTCAGCCTCATCCGGACCTGGGAGCTGTGCCAGTGTGGCAGCAGAA 701
 201 AsnGlyAspIleValGluGluGlnLysArgLeuLeuGlyLysIle 217
 702 AATGGGAGCATCGTGGAGGAGGAACAGAGCGCTGTGGAGCAAGCAT 751
 217 eThrGlyProGluGlyHisValLeuSerHisProGluValGluAlaG 234
 752 CACTGGTCTCAGGCGCATGTGCTCAGCCACCCAGAGAGTAGAGCCG 801
 234 luAlaValTyrArgAlaValThrIleAlaLysGlnAlaAsnCysProLeu 250
 802 AGGCTGTGTACAGAGCAGTCCACCTTGGCAAGCAGCCCACTGCCACTA 851
 251 TyrValThrLysValMetSerLysGlyAlaAlaAspAlaIleAlaGln 267
 852 TAGTCAACCAAGGTGATGAGCAAGGTGCAGCTGACATGGTTGCCAAGC 901
 267 aLysArgArgGlyValValValPheGlyGluProIleThrAlaSerLeu 284
 902 CAA3CGCAGGGGGTGTGCTCTTTGGGAACCTATCACTGCCAGCTGG 951
 284 lyThrAspGlySerHisTyrTrpSerLysAsnTrpAlaLysAlaAla 300
 952 GCACATGATGGCTCACACTACTGGAGCAAGAACTGGGCCAAGGCTGCAG 1001
 301 PheValThrSerProProValAsnProAspProThrThrAlaAspHisLe 317
 1002 TTGTCTCACTTACCCCTTATACCCCGGACCTACTACTGCAGACCACT 1051
 317 uThrCysLeuLeuSerSerGlyAspLeuGlnValThrGlySerAlaHisC 334
 1052 CACCTCTCTGCTGCTCCAGTGGGACCTCCAGGTGACAGGAGTGGCCACT 1101
 334 ysThrPheThrThrAlaGlnLysAlaValGlyLysAspAsnPheAlaLeu 350
 1102 GCACCTCTCACTTACGCCAGAGGCTGTGGCAAGACAACTTCCACTG 1151
 351 IleProGluGlyThrAsnGlyIleGluGluArgMetSerMetValTrpG 367
 1152 ATCCCGGAGGTAGTCAACGGTATAGAGAGGCGCATGCTGTGGTGGGA 1201
 367 uLysCysValAlaSerGlyLysMetAspGluAsnGluPheValAlaVal 384
 1202 GAAATGTGTGGCTTCAGGAGAAATGGACGAGATGAGTTCGTCGCCG 1251
 384 hrSerThrAsnAlaAlaLysIlePheAsnPheTyrProArgLysGlyArg 400

```

|||||
1252 CCAGCACAATGCTGCCAAATCTTCAATTTTACCCAGGAAGGGCGT 1301
401 ValAlaValGlySerAspAlaAspLeuValIleTrpAsnProLysAlaIle 417
|||||
1302 GTGGCGGTGGGCTCTGATGCTGACCTGCTATCTGGAACCCAGGGCCAC 1351
417 rLysIleIleSerAlaLysThrHisAsnLeuAsnValGluTrpAsnIleP 434
|||||
1352 GAAAGTCACTCTGCCAAGAGCCATAACCTGAATGTAGAGTACAACATCT 1401
434 heGluGlyValGluCysArgGlyAlaProAlaValValIleSerGlnGly 450
|||||
1402 TTGAAGGAGTGGAGTGGCGAGGAGTGGCCACGGTGGTCAATAGTCAGGGC 1451
451 ArgValAlaLeuGluAspGlyLysMetPheValThrProGlyAlaGlyAr 467
|||||
1452 AGAGTGGTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1501
467 gPheValProArgLysThrPheProAspPheValThrLysArgIleLysA 484
|||||
1502 CTTTCATCCCGGAGAGGAGTTCGGGAGCTTGTCTATAGAGGATAAAGG 1551
484 laArgAsnArgLeuAlaGluIleHisGlyValProArgGlyLeuTrpAsp 500
|||||
1552 CTCGCAACAGGCTAGCAGAGATCCACGGTGTGCTCGAGGCTGTACGAC 1601
501 GlyProValHisGluValMetValProAlaLysProGlySerGlyAlaPr 517
|||||
1602 GGGCTGTGCATGATGATGTTACTTGCAGGAGGAGGAGGAGGAGGAGGAG 1651
517 oAlaArgAlaSerCysProGlyLysIleSerValProValArgAsnL 534
|||||
1652 GCGCGTGCATCTCTTCAGGCAAGATCTCAGTCCACCCCGTGGCAACC 1701
534 euHisGlnSerGlyPheSerLeuSerGlySerGlnAlaAspAspHisIle 550
|||||
1702 TGACACAGTGGGGTTCAGCTACTGTGCTCTCAGGCTGAGATCACAT 1751
551 AlaArgArgThrAlaGlnLysIleMetAlaProProGlyGlyArgSerAs 567
|||||
1752 GCCAGACGTACGGCTCAGAGATCATGCAACCCCGGAGGAGGAGGAGGAG 1801
567 nIleThrSerLeuSer 572
|||||
1802 CATCAGCTCTCTTCC 1817

```

seq_name: /STDSB/gcgdata/geneseq/geneseq/NA1998.DAT:AAV60815

seq_documentation_block:

ID AAV60815 standard; cDNA; 1817 BP.

XX AC AAV60815;

XX DE 08-DEC-1998 (first entry)

XX DE Mouse ULIP-1 coding sequence.

XX DE Mouse; ULIP-1 coding sequence.

KW Mouse; ULIP-1 coding sequence.

KW neurodegenerative disorder; diagnosis; ss.

XX OS Mus musculus.

XX FH key

XX FT Location/Qualifiers

XX FT 23..1741

XX FT /*tag= a

XX FT /product= "ULIP-1"

XX FT /note= "Unc-33-like phosphoprotein 1"

XX FR2759701-AL.

XX PN 21-AUG-1998.

XX PD

XX XX

PF 19-FEB-1997; 97FR-0001961.
 XX
 PR 19-FEB-1997; 97FR-0001961.
 XX
 PA (INRM) INSM INST NAT SANTE & RECH MEDICALE.
 XX
 PI Aguera M, Belin MF, Byk T, Honnorat J, Kolattukudy P;
 PI Quach TT, Sobel A;
 XX
 DR WPI: 1998-449610/39.
 DR P-PSDB; AAW68486.
 XX
 PT Mouse and human ULIP polyptide(s) - useful in detection of
 PT para-neoplastic neurological syndromes
 XX
 PS Claim 3; Fig 9; 90pp; French.

CC This sequence represents the mouse Unc-33-like phospho-protein (ULIP)-1
 CC coding sequence. The sequence was isolated based on similarity to the
 CC rat ULIP sequence. Proteins of the ULIP family or their corresponding
 CC nucleic acids can be used in compositions for treating neurodegenerative
 CC disorders and neoplasms, especially for para-neoplastic neurological
 CC syndromes and/or for the early diagnosis of tumorigenesis.

SQ Sequence 1817 BP; 426 A; 506 C; 511 G; 374 T; 0 other;

alignment_scores:

Quality: 2325.00

Ratio: 4.429

Length: 571

Gaps: 0

Percent Similarity: 91.944

Percent Identity: 75.306

alignment_block:

US-09-367-496-8 x AAV60815

Align seg 1/1 to: AAV60815 from: 1 to: 1817

1 MetSerPheGlnGlyLysLysSerIleProArgIleThrSerAspArgLe 17
 |||||
 23 ATGCTCTATCAGGGGAGAAATAATATTCACCCATCAGGAGGATCGTCT 72
 |||||
 17 uleulleArgGlyGlyArgIleValAsnAspAspGlnSerPheTyrAlaA 34
 |||||
 73 TCTGATCAAAAGGTGGCAAGATGTGAATGATGACCATCTCTTAAGCAG 122
 |||||
 34 spValHisValGluAspGlyLeuIleLysGlnIleGlyGluAsnLeuIle 50
 |||||
 123 ACATATACATGGAAGATGGTGTGATCAAGCAAAATAGGAGAAACCTGATT 172
 |||||
 51 ValProGlyGlyIleHisThrIleAspAlaHisGlyLeuMetValLeuPr 67
 |||||
 173 GTACAGGAGGGGTGAAGACCATCGAGCCCACTCCAGATGGTGATTC 222
 |||||
 67 oGlyGlyValAspValHisThrArgLeuGlnMetProValLeuGlyMet 84
 |||||
 223 CGGAGGAATTCACCTGATCTACTCGCTTCCAGATGCCTGACCAAGGAATGA 272
 |||||
 84 hrProAlaAspAspPheCysGlnGlyThrLysAlaAlaLeuAlaGlyGly 100
 |||||
 273 CATCCGCTGATGACTTCTTCAGGGAACCAAGCGCCCTGGCGGGGGA 322
 |||||
 101 ThrThrMetIleLeuAspHisValPheProAspThrGlyValSerLeuLe 117
 |||||
 323 ACCACCATGATCATTCACCATGTTCTTCCAGCCCGGAGGAGGAGGAGGAG 372
 |||||
 117 uAlaAlaTyrGluGlnTrpArgGluArgAlaAspSerAlaAlaCysCysA 134
 |||||
 373 GGCTGCTTGTGATCAGTGGAGGAGTGGGCTGACAGCAAGTCTGCTGTG 422
 |||||
 134 spTyrSerLeuHisValAspIleThrArgTrpHisGluSerIleLysGlu 150
 |||||
 423 ACTATTTCGCTGACCGTGGACATCATGTAGTGGCAACAGGAGGAGGAGGAG 472

XX

CC This sequence represents the mouse Unc-33-like phospho-protein (ULIP)-3
 CC coding sequence. The sequence was isolated based on similarity to the
 CC rat ULIP sequence. Proteins of the ULIP family or their corresponding
 CC nucleic acids can be used in compositions for treating neurodegenerative
 CC disorders and neoplasms, especially for para-neoplastic neurological
 CC syndromes and/or for the early diagnosis of tumorigenesis.

XX Sequence 2297 BP; 511 A; 653 C; 632 G; 501 T; 0 other;
 SQ

alignment_scores:

Quality: 2151.00 Length: 571
 Ratio: 4.226 Gaps: 0
 Percent Similarity: 89.142 Percent Identity: 69.002

alignment_block:

US-09-367-496-8 x AAV60816 ..

Align seg 1/1 to: AAV60816 from: 1 to: 2297

1 MetSerPheGlnGlyLysLysSerIleProArgIleThrSerAspArgLe 17
 270 ATGTCCTCATCAGGGAAGAAGACATCCCGACATCACCAGTGAACGGCT 319
 17 uLeuIleArgGlyGlyArgIleValAsnAspAspGlnSerPheTyAlaA 34
 320 CCTCATCAGAGTGGACGATCATCATGATGACCATCTCTTCTAGCGCG 369
 34 spValHisValGluAspGlyLeuIleGlyGlnIleGlyGluAsnLeuIle 50
 370 ATGTCCTACCTAGAGTGGACGATCATCATGATGACCATCTCTTCTAGCGCG 369
 51 ValProGlyGlyIleHisThrIleAspAlaHisGlyLeuMetValLeuPr 67
 420 GTTCTGTGGAGTGAAGACCATCAGGCGAATGCCCAATGGTGTATCC 469
 67 OGlyGlyValAspValHisThrArgLeuGlnMetProValLeuGlyMet 84
 470 CGTGGCATGATGATCAACATCTTACCTGACAGGCGCTCCAGGCGATGA 519
 84 hrProAlaAspPheCysGlnGlyThrLysAlaAlaLeuAlaGlyGly 100
 520 CCGTGGCTGATGACTTCTTCAGGCGCTAAAGCAGCGCTGGCAGGTGA 569
 101 ThrThrMetLeuLeuAspHisValPheProAspThrGlyValSerLeuLe 117
 570 ACCAGATGATCATGACCGTGTGTTCTCAACCTGGTCCAGCTGTGT 619
 117 uAlaAlaTyrgluGlnTrpArgGluArgAlaAspSerAlaAlaCysCys 134
 620 GACTTCCTTTGAGAAATGGCAGGACGAGCAGACACCAATCTGCTGTG 669
 134 spTySerLeuHisValAspIleThrArgTrpHisGluSerIleLysGlu 150
 670 ACTATTCCTCCAGTGGACATCAGACGCTGTATGATGTTGTCGGAA 719
 151 GluLeuAlaLeuValLysGluLysGlyValAsnSerPheLeuValPh 167
 720 GAGTGGAGTGTGTGGAGGACAAAGGTGTCACTCTTCCAGTCTA 769
 167 eMetAlaTyrgluLysAspArgCysGlnCysSerAspSerGlnMetTyrglu 184
 770 CATGGGTATAGGACCTCTACCATGATGCTGACAGCCAGCTGTATGA 819
 184 IePheSerIleIleArgAspLeuGlyAlaLeuAlaGlnValHisAlaGlu 200
 820 CCTTACCTCTCTTAAAGGTTTGGAGCTGTGATCTTGTAGTCCATGCA 869
 201 AsnGlyAspIleValGluGluGlnLysArgLeuLeuGluLeuGlyIle 217
 870 AATGAGATTGTAGTACGAGACAAACGATGCTGTGGAGTGGCAT 919

217 eThrGlyProGluGlyHisValLeuSerHisProGluValGluAlaG 234
 920 CACGGGTCCTCCAGGCTCATGCTGACGACACCGAGGAGCTGGAGCCG 969
 234 luAlaValTyrgluAlaValThrIleAlaLysGlnAlaAsnCysProLeu 250
 970 AGCTGTGTTCCGGGCTATTGCCATTGCGCGGATCAATTGCCCTGTG 1019
 251 TyrValThrLysValMetSerLysGlyAlaAlaAspAlaIleAlaGln 267
 1020 TACATCACCAAGGTCTATGACGAGAGTGCAGCGACATCATCGCTGC 1069
 267 aLysArgArgGlyValValPheGlyGluPheIleThrAlaSerLeuG 284
 1070 CAGGAAGAAGGCCCTCTTGTCTTCGTTGAGCCATAGCCGCGAGCTGG 1119
 284 LysThrAspGlySerHisTyrrTrpSerLysAsnTrpAlaLysAlaAla 300
 1120 GAACCGATGGCACCCACTACTGGAGCAAGACTGGGCCAAGCAGCTGCA 1169
 301 PheValThrSerProValAsnProAspProThrThrAlaAspHisLe 317
 1170 TTTGTGACTTCCCTCCCTGAGCCAGACCCACCACCTCCTGACTACT 1219
 317 uThrCysLeuLeuSerSerGlyAspLeuGlnValThrGlySerAlaHisC 334
 1220 GACCTCTTGTGCTGCTGCTGAGACTTTCAGGTCCACAGTGTGCTGCT 1269
 334 ysThrPheThrThrAlaGlnLysAlaValGlyLysAspAsnPheAlaLeu 350
 1270 GTCCCTACAGTATTGCTCAGAAAGCTGTGGCAAGAGGACAACTTCACTCTG 1319
 351 IleProGluGlyThrAsnGlyIleGluGluArgMetSerMetValTrpG 367
 1320 ATCCCTGAGGGTGTCAATGTATAGAGAGCGGATGACCGTGTCTGGGA 1369
 367 uLysCysValAlaSerGlyLysMetAspGluAsnGluPheValAlaVal 384
 1370 CAAGGCGATGGCTACTGGCAAGATGGATGAGAACCAAGTTGTAGCCGTCA 1419
 384 hrSerThrAsnAlaAlaLysIlePheAsnPhetTyrrProArgLysGlyArg 400
 1420 CCAGCACCAACGAGCAAGATCTTCAACCTGTACCCGAGGAAAGTCCG 1469
 401 ValAlaValGlySerAspAlaAspLeuValIleTrpAsnProLysAla 417
 1470 ATCGCTGTGGCTCCGATGCTGACGTAGTCACTGGGACCCAGATAAGAT 1519
 417 rLysIleIleSerAlaLysThrHisAsnLeuAsnValGluTyrrAsnIle 434
 1520 GAAGACCAATACAGCAAAAGCCATAAATCAACTGTGGAGTACACATCT 1569
 434 heGluGlyValGluCysArgGlyLysMetPheValThrProGlyAlaGly 450
 1570 TTGAGGCGATGGAGTGGCCACGGCTCCCTGGTGTGTCATCAGTCAGGCG 1619
 451 ArgValAlaLeuGluAspGlyLysMetPheValThrProGlyAlaGly 467
 1620 AAGATTGTCTTGGAGTGGAAACATCAGTGTGTCAGCAAGGCGATGGCCG 1669
 467 gPheValProArgLysThrPheProAspPheValTyrrLysArgIleLys 484
 1670 CTTCATCCCTCGAAGCCATTCAGAGGATCTCTACCGGCTGTGAGCA 1719
 484 laArgAsnArgLeuAlaGluIleHisGlyValProArgGlyLeuTyrrAsp 500
 1720 TCAGAGCAAGGTTTTCGGGTGTGATAGTGTTCAGGCGGATGTACGAT 1769
 501 GlyProValHisGluValMetValProAlaLysProGlySerGlyAlaPr 517
 1770 GGCCTGTGTACGAGGTGCGAGCTACACCAACATGCTGCTCTGCTCC 1819
 517 oAlaArgAlaSerCysProGlyLysIleSerValProValArgAsnL 534

[illegible]

PR	05-AUG-1999;	99US-0147192;
PR	05-AUG-1999;	99US-0147260;
PR	06-AUG-1999;	99US-0147303;
PR	06-AUG-1999;	99US-0147436;
PR	09-AUG-1999;	99US-0147493;
PR	10-AUG-1999;	99US-0147935;
PR	11-AUG-1999;	99US-0148171;
PR	12-AUG-1999;	99US-0148319;
PR	13-AUG-1999;	99US-0148341;
PR	13-AUG-1999;	99US-0148565;
PR	16-AUG-1999;	99US-0148684;
PR	17-AUG-1999;	99US-0149368;
PR	18-AUG-1999;	99US-0149175;
PR	20-AUG-1999;	99US-0149426;
PR	20-AUG-1999;	99US-0149722;
PR	20-AUG-1999;	99US-0149723;
PR	23-AUG-1999;	99US-0149929;
PR	23-AUG-1999;	99US-0149902;
PR	25-AUG-1999;	99US-0149930;
PR	26-AUG-1999;	99US-0150566;
PR	27-AUG-1999;	99US-0150884;
PR	27-AUG-1999;	99US-0151065;
PR	27-AUG-1999;	99US-0151066;
PR	27-AUG-1999;	99US-0151080;
PR	30-AUG-1999;	99US-0151303;
PR	31-AUG-1999;	99US-0151348;
PR	01-SEP-1999;	99US-0151930;
PR	07-SEP-1999;	99US-0152363;
PR	10-SEP-1999;	99US-0153070;
PR	13-SEP-1999;	99US-0153758;
PR	15-SEP-1999;	99US-0154018;
PR	16-SEP-1999;	99US-0154039;
PR	20-SEP-1999;	99US-0154779;
PR	22-SEP-1999;	99US-0155139;
PR	23-SEP-1999;	99US-0155139;
PR	24-SEP-1999;	99US-0155486;
PR	28-SEP-1999;	99US-0155659;
PR	29-SEP-1999;	99US-0156458;
PR	04-OCT-1999;	99US-0156596;
PR	05-OCT-1999;	99US-0157117;
PR	06-OCT-1999;	99US-0157753;
PR	07-OCT-1999;	99US-0157865;
PR	08-OCT-1999;	99US-0158029;
PR	12-OCT-1999;	99US-0158323;
PR	13-OCT-1999;	99US-0158369;
PR	13-OCT-1999;	99US-0159293;
PR	13-OCT-1999;	99US-0159294;
PR	14-OCT-1999;	99US-0159295;
PR	14-OCT-1999;	99US-0159329;
PR	14-OCT-1999;	99US-0159330;
PR	14-OCT-1999;	99US-0159331;
PR	14-OCT-1999;	99US-0159637;
PR	18-OCT-1999;	99US-0159638;
PR	21-OCT-1999;	99US-0159584;
PR	21-OCT-1999;	99US-0160741;
PR	21-OCT-1999;	99US-0160767;
PR	21-OCT-1999;	99US-0160768;
PR	21-OCT-1999;	99US-0160770;
PR	22-OCT-1999;	99US-0160814;
PR	22-OCT-1999;	99US-0160815;
PR	22-OCT-1999;	99US-0160980;
PR	22-OCT-1999;	99US-0160981;
PR	25-OCT-1999;	99US-0161009;
PR	25-OCT-1999;	99US-0161404;
PR	25-OCT-1999;	99US-0161405;
PR	26-OCT-1999;	99US-0161406;
PR	26-OCT-1999;	99US-0161359;
PR	26-OCT-1999;	99US-0161360;
PR	28-OCT-1999;	99US-0161361;
PR	28-OCT-1999;	99US-0161920;
PR	28-OCT-1999;	99US-0161992;
PR	29-OCT-1999;	99US-0161993;
PR	29-OCT-1999;	99US-0162142;

us-09-367-496-8.std.rng

Tue Jul 31 13:08:18 2001

```

548 p 54E
1788 T 1788

seq_name: /SIDBS/gcgdata/geneseq/geneseq/NA2000.DAT:AAC40092
seq_documentation_block:
ID AAC40092 standard; DNA; 1734 BP.
XX
AC AAC40092;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 27016.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EPI0334C5-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 04-MAY-1999; 99US-0132048.
PR 05-MAY-1999; 99US-0132407.
PR 06-MAY-1999; 99US-0132484.
PR 07-MAY-1999; 99US-0132485.
PR 11-MAY-1999; 99US-0132486.
PR 14-MAY-1999; 99US-0132487.
PR 14-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.

```

PR 18-JUN-1999; 99US-0139454.
 PR 18-JUN-1999; 99US-0139455.
 PR 18-JUN-1999; 99US-0139456.
 PR 18-JUN-1999; 99US-0139457.
 PR 18-JUN-1999; 99US-0139458.
 PR 18-JUN-1999; 99US-0139459.
 PR 18-JUN-1999; 99US-0139460.
 PR 18-JUN-1999; 99US-0139461.
 PR 18-JUN-1999; 99US-0139462.
 PR 18-JUN-1999; 99US-0139463.
 PR 18-JUN-1999; 99US-0139750.
 PR 21-JUN-1999; 99US-0139763.
 PR 22-JUN-1999; 99US-0139817.
 PR 23-JUN-1999; 99US-0139899.
 PR 23-JUN-1999; 99US-0140353.
 PR 24-JUN-1999; 99US-0140354.
 PR 28-JUN-1999; 99US-0140695.
 PR 29-JUN-1999; 99US-0140823.
 PR 30-JUN-1999; 99US-0140991.
 PR 01-JUL-1999; 99US-0141287.
 PR 02-JUL-1999; 99US-0141844.
 PR 06-JUL-1999; 99US-0142154.
 PR 08-JUL-1999; 99US-0142055.
 PR 09-JUL-1999; 99US-0142390.
 PR 12-JUL-1999; 99US-0142803.
 PR 13-JUL-1999; 99US-0142977.
 PR 14-JUL-1999; 99US-0142920.
 PR 15-JUL-1999; 99US-0143542.
 PR 16-JUL-1999; 99US-0143624.
 PR 16-JUL-1999; 99US-0144005.
 PR 19-JUL-1999; 99US-0144085.
 PR 19-JUL-1999; 99US-0144086.
 PR 19-JUL-1999; 99US-0144325.
 PR 19-JUL-1999; 99US-0144331.
 PR 19-JUL-1999; 99US-0144332.
 PR 19-JUL-1999; 99US-0144333.
 PR 19-JUL-1999; 99US-0144334.
 PR 20-JUL-1999; 99US-0144335.
 PR 20-JUL-1999; 99US-0144352.
 PR 20-JUL-1999; 99US-0144632.
 PR 21-JUL-1999; 99US-0144814.
 PR 21-JUL-1999; 99US-0145086.
 PR 21-JUL-1999; 99US-0145088.
 PR 22-JUL-1999; 99US-0145088.
 PR 22-JUL-1999; 99US-0145087.
 PR 22-JUL-1999; 99US-0145089.
 PR 22-JUL-1999; 99US-0145192.
 PR 23-JUL-1999; 99US-0145145.
 PR 23-JUL-1999; 99US-0145218.
 PR 26-JUL-1999; 99US-0145224.
 PR 27-JUL-1999; 99US-0145276.
 PR 27-JUL-1999; 99US-0145913.
 PR 27-JUL-1999; 99US-0145918.
 PR 28-JUL-1999; 99US-0145919.
 PR 02-AUG-1999; 99US-0145951.
 PR 02-AUG-1999; 99US-0146386.
 PR 02-AUG-1999; 99US-0146388.
 PR 03-AUG-1999; 99US-0146389.
 PR 04-AUG-1999; 99US-0147038.
 PR 04-AUG-1999; 99US-0147204.
 PR 05-AUG-1999; 99US-0147302.
 PR 05-AUG-1999; 99US-0147192.
 PR 06-AUG-1999; 99US-0147260.
 PR 06-AUG-1999; 99US-0147303.
 PR 09-AUG-1999; 99US-0147416.
 PR 09-AUG-1999; 99US-0147493.
 PR 10-AUG-1999; 99US-0147935.
 PR 11-AUG-1999; 99US-0148171.
 PR 12-AUG-1999; 99US-0148319.
 PR 13-AUG-1999; 99US-0148341.
 PR 13-AUG-1999; 99US-0148565.
 PR 13-AUG-1999; 99US-0148684.

PR 16-AUG-1999; 99US-0149368.
 PR 17-AUG-1999; 99US-0149175.
 PR 18-AUG-1999; 99US-0149426.
 PR 20-AUG-1999; 99US-0149722.
 PR 20-AUG-1999; 99US-0149723.
 PR 20-AUG-1999; 99US-0149929.
 PR 23-AUG-1999; 99US-0149902.
 PR 23-AUG-1999; 99US-0149930.
 PR 25-AUG-1999; 99US-0150566.
 PR 26-AUG-1999; 99US-0150884.
 PR 27-AUG-1999; 99US-0151065.
 PR 27-AUG-1999; 99US-0151066.
 PR 30-AUG-1999; 99US-0151080.
 PR 31-AUG-1999; 99US-0151303.
 PR 01-SEP-1999; 99US-0151438.
 PR 07-SEP-1999; 99US-0151930.
 PR 10-SEP-1999; 99US-0152363.
 PR 13-SEP-1999; 99US-0153070.
 PR 15-SEP-1999; 99US-0153758.
 PR 16-SEP-1999; 99US-0154018.
 PR 20-SEP-1999; 99US-0154039.
 PR 22-SEP-1999; 99US-0154779.
 PR 23-SEP-1999; 99US-0155139.
 PR 24-SEP-1999; 99US-0155486.
 PR 28-SEP-1999; 99US-0155659.
 PR 29-SEP-1999; 99US-0156458.
 PR 29-SEP-1999; 99US-0156596.
 PR 04-OCT-1999; 99US-0157117.
 PR 05-OCT-1999; 99US-0157753.
 PR 06-OCT-1999; 99US-0157865.
 PR 07-OCT-1999; 99US-0158029.
 PR 12-OCT-1999; 99US-0158232.
 PR 13-OCT-1999; 99US-0158369.
 PR 13-OCT-1999; 99US-0159293.
 PR 14-OCT-1999; 99US-0159294.
 PR 14-OCT-1999; 99US-0159295.
 PR 14-OCT-1999; 99US-0159329.
 PR 14-OCT-1999; 99US-0159330.
 PR 14-OCT-1999; 99US-0159331.
 PR 14-OCT-1999; 99US-0159637.
 PR 18-OCT-1999; 99US-0159638.
 PR 21-OCT-1999; 99US-0159584.
 PR 21-OCT-1999; 99US-0160741.
 PR 21-OCT-1999; 99US-0160767.
 PR 21-OCT-1999; 99US-0160768.
 PR 21-OCT-1999; 99US-0160770.
 PR 21-OCT-1999; 99US-0160814.
 PR 21-OCT-1999; 99US-0160815.
 PR 22-OCT-1999; 99US-0160980.
 PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160989.
 PR 25-OCT-1999; 99US-0161404.
 PR 25-OCT-1999; 99US-0161405.
 PR 26-OCT-1999; 99US-0161406.
 PR 26-OCT-1999; 99US-0161359.
 PR 26-OCT-1999; 99US-0161360.
 PR 28-OCT-1999; 99US-0161361.
 PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161992.
 PR 28-OCT-1999; 99US-0161993.
 PR 29-OCT-1999; 99US-0162142.

alignment_scores:

Quality: 1018.00
 Ratio: 2.844
 Percent Similarity: 68.451
 Length: 523
 Gaps: 9
 Percent Identity: 41.491

alignment_block:

US-09-367-496-8 x AAC40092

Align seg 1/1 to: AAC40092 from: 1 to: 1734

12 IleThrSerAspArgLeuIleArgGlyGlyArgIleValAsnAspAs 28

Tue Jul 31 13:08:18 2001

```

1018 .....CTAGGACATGGGAAAGCCCTACAAGATGCCCTTTCCACAGGA 1059
325 AspLeuGlnValThrGlySerAlaHisCysThrPheThrAlaGlnLy 341
1060 ATCTTCAGCTTGTAGGAATGATCATCTGCTCAATTTCAATTTACACAAA 1109
341 sAlaValGlyLysAspAsnPheAlaLeuIleProGluGlyThrAsnGlyI 358
1110 AGCTCTAGGACTTGTATGTTTCGCAAAATACCTAATGTTGTTAATGGCC 1159
358 leGluGluArgMetSerMetValTrpGluLysCysValAlaSerGlyLys 374
1160 TTGAGGACCGATGCTACTTCATATGGACACAGCATGCTGGAGTCTGCCAA 1209
375 MetaspGluAsnGluPheValAlaValThrSerThrAsnAlaLysI 391
1210 CTCTCAGCTACTGATTAATGTTTCAATTAACCAAGCATCTGCTGGTCTAGAA 1259
391 ePheAsnPheThrProArgLysGlyArgValAlaValGlySerAlaLysThr 424
1260 TTTCACATATATCCACGAAAGGAGCTATCTTCTGCTGGTCTGGATGAG 1309
408 sPLeuValIleTrpAsnProLysAlaThrLysIleIleSerAlaLysThr 424
1310 ATATATCATATTTGAATCAACCAAGCTACGAGATTAGCTCAAGTCT 1359
425 HisAsnLeuAsnValGluTyAsnIlePheGluGlyValGluCysArgL 441
1360 CATATTCAGATTCAGACACAAAGCTCTACGAGGCGCAGAGGAAAGGG 1409
441 yAlaProAlaValIleSerGlnGlyArgValAlaLeuGluaspGlyL 458
1410 AAAAGTTCAAGTGACAATAGCAGGAGGAGCAATTTGTGGGAAACACGAG 1459
458 ysMetPheValThrProGlyAlaGlyArgPheValProArgLysThrPhe 474
1460 AACTTAAAGTTGTTCCCAAGAGTGCAAGTATATA.....GAGATG 1500
475 ProAspPheValTyLysArgIleLysAlaArgAsnArgLeuAlaGluI 491
1501 CCTCTTTCAGTTAC..... 1515
491 eHisGlyValProArgGlyLeuTyAspGly..... 501
1516 .....CTTTTCATGCTATTGAGAAATCAGATGCTA 1546
502 .....ProValHisGluValMetValProAla 510
1547 ATTATCTATCTCTCTTCGAGCTCCAGCTTAAGCGTGTGCAAACTGAAGCT 1596
511 LysProGlySerGly 515
1597 ACCTAAAGTGCAGGT 1611

seq_name: /SIDS8/gcgdata/geneseq/geneseq/NA1996.DAT:AAAT31258
seq_documentation_block:
ID AAT31258 standard; DNA; 1569 BP.
XX
AC AAT31258;
XX
DT 31-JAN-1997 (first entry)
XX
DE Hydatoninase coding sequence.
XX
KW Hydatoninase; hydantoin; d-N-carbamoyl-alpha-amino acid;
KW intermediate; drug synthesis; penicillin; cephalosporin;
KW antibiotic; ss.
XX
OS Pseudomonas sp.
XX
FH Key
FT CDS 60..1518

```


[illegible]

```

||||| ..... :||| ||||| |||||
1251 AAGAAATCATCTGTTCCGTGCAGAAATCCATCATCAATAATCTT 1300

      eGluglyValGluCysArgGlyAlaProAlaValIleSerGlnGlyA 451
||||| ||| ||| ||| :||| ||||| :|||
1301 CGAGGGCTACGAGGTGACCGCATGCCGCGATACAGGCTTCACGCGCG 1350

      rglValAlaLeu... GluaspgLyLysMetPheValThrProGlyAlaGly 466
:||||:: |||:::|||::: ||||| |||
1351 AGGTGCTATGCGGCGAGGAAGCGCACGAAACCAGCGCGCCGCGT 1400

      ArgPheValProArgLyThrPhe 474
|||||| |||::: |||
1401 AGGTCGTCGAGCGCGGCCCTTT 1424

seq_name: /SID88/gcdata/geneseq/geneseqn/NA1992.DAT:AAQ31987

seq_documentation_block:
ID AAQ31987 standard; DNA; 1938 BP.
XX
XX AAQ31987;
XX
DT 05-MAY-1993 (first entry)
XX
DE Heat stable hydantoinase gene.
XX
KW Hydantoinase; E.coli; heat; stable; thermal; ss.
XX
OS Bacillus stearothermophilus.
XX
XX Key Location/Qualifiers
FH CDS 345..1760
FT FT /*tag= a
FT FT
XX JP04325093-A.
PN
PD 13-NOV-1992.
XX
XX XX 23-APR-1991; 91JP-0117802.
XX
XX XX 23-APR-1991; 91JP-0117802.
XX
XX (NIPS ) NIPPON SODA CO.
PA
XX WPI; 1992-428828/52.
XX P-PSDB; AAR29715.
DR
XX
XX Heat-stable hydantoinase (I) gene - has specified base and
PT aminoacid sequence and is prepd. from E. Coli
PT Claim 1; Page 5-7; 8pp; Japanese.
PS
XX This sequence encodes a heat-stable hydantoinase protein. This
CC sequence can be used to transform E.coli such that the protein is
CC produced stably by the microbe.
XX Sequence 1938 BP; 582 A; 342 C; 530 G; 484 T; 0 other;
SQ

alignment_scores:
Quality: 837.00 Length: 458
Ratio: 2.607 Gaps: 4
Percent Similarity: 70.087 Percent Identity: 39.956

alignment_block:
US-09-367-496-8 x AAQ31987 ..
Align seg 1/1 to: AAQ31987 from: 1 to: 1938

18 LeutlleArgGlyGlyArgIleValAsnAspGlnSerPheTyraLaas 34
||||| :::::: ||||| :::: |||||
354 TTGTAATAAATTGAACATGTTCACCGCTACAGATATATATGAGCGGA 403
```

```

1295 CTCAATTTGATTTAAAGCCAAAAGATATTAGGAAGGGAGATTTTACCA 1344
350 euileProGluGlyThrAsnGlyIleGluGluArgMetSerMetValTrp 366
1345 AAATCCCAATATGGTGCTTATTATGAGGATCGGGTGGATTTCTTTTC 1394
367 GluLysCysValAlaSerGlyLysMetAspGluAsnGluPheValAlaVa 383
1395 AGTGAAGGAGTGAAAGGAGAGATATACCTCAACCAAGTTTGTGTATAT 1444
383 lThrSerThrAsnAlaAlaLysIlePheAsnPheTyProArgLysGlyA 400
1445 TGTATCAACAAGATCCCAAAATTTGTTGGTCTATTCCCGAAGAAAGGAA 1494
400 rgValAlaValGlySerAspAlaAspLeuValIleTrpAsnProLysAla 416
1495 CCATTCGCGTGGTGGGATGCGGATTTAGTCATTTTGTATCCCAACGGT 1544
417 ThrLysIleSerAlaLysThrHisAsnLeuAsnValGluTyAsnIle 433
1545 GAACGGGTGATTTTCAGCCGAACACACCATATGCGTGTGGATTATATCC 1594
433 ePheGluGlyValGluCysArgGlyAlaProAlaValValIleSerClnG 450
1595 GTTTGAAGGATGAAGATACAGGGGAACCTGTGCGTTTATGTAGAG 1644
450 lyArgValAlaLeuGluAspGlyLysMetPheValThrProGlyValAlaGly 466
1645 GAGAAATTTGTGTACGTGATATAAATTTTCGCGAAGACCGGGGTACGGC 1694
467 ArgPheValProArgLysThrPhe 474
1695 CAATATGTTAAACGCGCGAAATAT 1718

```

seq_name: /SID88/gcgdata/geneseq/geneseq/NA1995.DAT:AAQ80565

seq_documentation_block:

ID AAQ80565 standard; DNA; 1383 BP.

AC AAQ80565;

DT 27-OCT-1995 (first entry)

DE B.thermoglucoasius D-hydantoinase gene with engineered mutation.

KW D-hydantoinase; dihydropyrimidinase; mutein; variant; mutation;

KW D-N-carbamoyl-alpha-amino acid; penicillin; cephalosporin;

KW production; racemisation; ss.

OS Bacillus thermoglucoasius.

Key Location/Qualifiers

FT 1..1383

FT /*tag= a

FT /EC_number= 3.5.2.2

FT /note="Deletion of a single base in the wild-type

FT ORF results in this sequence coding for an

FT enzyme which is 12 amino acids shorter than

FT the wild-type protein and the last 6 amino

FT acids are different"

XX DB4328829-A.

XX 02-MAR-1995.

XX 27-AUG-1993; 93DE-4328829.

XX 27-AUG-1993; 93DE-4328829.

XX (BOE) BOEHRINGER MANNHEIM GMBH.

XX Burtcher H, Lang G, Popp F;

XX

DR WPI; 1995-099331/14.

XX P-PSDB; AAR69874.

XX New highly active and stable D-hydantoinase enzyme - isolated

PT from Bacillus thermoglucoasius, useful in the prodn. of

PT D-N-carbamoyl-alpha-amino acid for synthesis of, e.g. penicillin

XX Claim 3; Page 7; 8pp; German.

XX The wild-type D-hydantoinase gene was isolated from Bacillus

CC thermoglucoasius using the amplification primers Hyd1 and Hyd2 (see

CC AAQ80566 and AAQ80567). A single-base deletion in the HindIII site (from

CC AAGCTT to AAGCT) was introduced to give a coding region (AAQ80565)

CC that encoded a D-hydantoinase (AAR69874) that was shorter than the

CC wild-type protein with a different C-terminal sequence (see Features

CC Table). The mutant enzyme was found to have improved heat stability

CC and higher activity compared to known hydantoinases.

XX Sequence 1383 BP; 427 A; 251 C; 368 G; 337 T; 0 other;

alignment_scores:

Quality: 826.00 Length: 458

Ratio: 2.614 Gaps: 4

Percent Similarity: 68.996 Percent Identity: 39.738

alignment_block:

US-09-367-496-8 x AAQ80565

Align seg 1/1 to: AAQ80565 from: 1 to: 1383

18 LeuIleArgGlyGlyArgIleValAsnAspGlnSerPheTyrAlaAs 34

10 ATATAATAAATGGAACGATTTTACCGCAACCGCATACGTATGAAGCGGA 59

34 pValHisValGluAspGlyLeuIleLysGlnIleGlyGluAsnLeuIleV 51

60 CTGTGCTCATTAAGACGGAAATTTGCCATGATAGCCCAACATTTA...G 106

51 alProGlyGlyIleHisThrIleAspAlaHisGlyLeuMetValLeuPro 67

107 AAGAAAAGCGCGTGAAGTATGATGATGCAAGGCTGTACGTATTTCCTCA 156

68 GlyGlyValAspValHisThrArgLeuGlnMetProValLeuGlyMetTh 84

157 GCGGTATTGATTCGCACACGATTTAGATATGCCCTTTGGCGGCGCGGT 206

84 rProAlaAspPheCysGlnGlyThrLysAlaAlaLeuAlaGlyGlyT 101

207 GACAAAGGATGATTTTCAATCTGGAACGATTTGCGCGGCGCATTTGGCGGA 256

101 hrThrMetIleLeuAspHisValPheProAspThrGlyValSerLeuLeu 117

257 CAACGACCATCATCGACTTTTGTAAACGAATAAAGGGAGGCATTAAAA 306

118 AlaAlaTyrGluGlnTrpArgGluArgAlaAspSerAlaAlaCysCysAs 134

307 AAAGCGATTGAACTTGGCACAACAAGCGAAGGAAAGCGGTATTTCGA 356

134 pTyrSerLeuHisValAspIleThrArgTrpHisGluSerIleLysGluG 151

357 TTATGGCTTCCATTTAATGATGAGGAAATTTACGGATGATGATGATGAG 406

151 luLeuGluAlaLeuValLysGluLysGlyValAsnSerPheLeuValPhe 167

407 AGCTGCCAAAAGTCATTCCCGAAGAGGAGGATAACATCTTTAAAGTGTT 456

168 MetaLalaTyrLysAspArgCysGlnCysSerAspSerGlnMetTyrGluI 184

457 ATGGCGTATAAAACGATTTTTCAGGCAGATGATGGAACGTTATATACCGCAC 506

184 ePheSerIleIleArgAspLeuGlyAlaLeuAlaGlnValHisAlaGluA 201

Tue Jul 31 13:08:18 2001

```
seq_documentation_block:
ID   AAN70154 standard; DNA; 1883 BP.
XX
AC   AAN70154;
XX
DT   07-APR-1991 (first entry)
XX
DE   Hydatonase gene contg. construct Lu 1220.
XX
KW   Construct Lu 1220; mesophilic microorganism; thermophilic hydantoinase;
XX   ss.
OS   CBS 303.80.
XX
PN   EP219034-A.
XX
PD   22-APR-1987.
XX
PR   09-OCT-1985; 85DE-3535987.
XX
PA   (BADI ) BASF AG.
XX
PI   Jacob E, Henco K, Marciniowski S, Schenk G;
XX
DR   WPI; 1987-110013/16.
XX
PT   Prodn. of mesophilic microorganisms contg. thermophilic hydantoinase -
XX   by genetic modification with DNA from thermophilic microorganisms.
XX
PS   Disclosure; Fig. 12A-B; 27pp; English.
XX
CC   The construct Lu 1220 contains a DNA sequence from a thermophilic
CC   microorganism which cleaves D-hydantoin. The DNA fragment encoding
CC   hydatonase comprises nucleotides 391-1746.
CC   The enzyme is useful for seroselective conversion of D,L-hydantoin to
CC   D,N-carbamoyl-alpha amino acids, which are intermediates for D-alpha
CC   amino acids. These are useful in the prodn. of penicillins and
CC   cephalosporins. Transformed mesophilic microorganisms, esp. E.coli,
CC   contg. the DNA can produce 4-40 times greater enzyme activity than
CC   wild type organisms. The DNA is from a thermophilic organism, esp.
CC   CBS 303.80 or 363.80. The cloning vector is linearised cosmid or
CC   lambda phage DNA.
CC   See also AAN70153.
XX
SQ   Sequence 1883 BP; 592 A; 340 C; 482 G; 469 T; 0 other;

alignment_scores:
Quality: 805.00 Length: 458
Ratio: 2.547 Gaps: 4
Percent Similarity: 68.996 Percent Identity: 39.738

alignment_block:
US-09-367-496-8 x AAN70154 ..
Align seg 1/1 to: AAN70154 from: 1 to: 1883
18 LeuileArgGlyGlyArgIleValAlaAspGlnSerPheTyrAlaAs 34
   ::::::::::::::::::::|
400 ATAATAAAATGCAACGATTGTTACCGCAACCGATACGTATGAAGCGGA 449
   ::::::::::::::::::::|
34 pValHisValGluAspGlyLeuIleLysGlnIleGlyGluAsnLeuIle 51
   ::::::::::::::::::::|
450 CTTCGTCATTAAAGACGGAATAATGGCATATAGGCCAACATTTA...G 496
   ::::::::::::::::::::|
51 alProGlyGlyIleHisThrIleAspAlaHisGlyLeuMetValLeuPro 67
   ::::::::::::::::::::|
497 AAGAAAAAGCGCTGAAGTGTGATGCAAAAGCTGTTACGTATTTCGA 546
   ::::::::::::::::::::|
68 GlyGlyValAspValHisThrArgLeuGlnMetProValLeuGlyMetTh 84
   ::::::::::::::::::::|
547 GCCGGTATTGATTCGCACACGATTTAGATATATGCGGTTTGGCGGCACGGT 596
```

seq_name: /SID8/gcgdata/geneseq/geneseq/NA1987.DAT: AAN70154

```

84 rProAlaAspPheCysGlnGlyThrLysAlaAlaLeuAlaGlyGlyT 101
: ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : |||||
597 GCAAAAGGATGATTCGAATCTGGAACGATTCGCGCGCATTTGGCGGAA 646
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
101 hrThrMetIleLeuAspHisValPheProAspThrGlyValSerLeuLeu 117
||||| ||||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
647 CAACGACCATCATCGACTTTCTTTTAAACGAATAAAGGGGAGCCATTAAAA 696
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
118 AlaAlaTyrGluGlnTyrArgGluAlaAspSerAlaAlaCysCysAs 134
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
697 AAAGCGATTGAACCTGGCACACAAAGCGAAGGAAAGCGTTATTGA 746
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
134 pTyrSerLeuHisValAspIleThrArgTyrHisGluSerIleLysGluG 151
||||| ||||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
747 TTATGGCTTCCATTTAATGATTAGCGAATACGATGACGTTATTAGA 796
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
151 LuLeuGluAlaLeuValLysGluLysGlyValAlaSerPheLeuValPhe 167
||||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
797 AGCTCCCAAGATCATTCGCAAGAGGGAATACATCTTTAAAGTGTTT 846
||||| ||||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
168 MetAlaTyrLysAspArgCysGlnCysSerAspSerGlnMetTyrGluI 184
||||| ||||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
847 ATGGCGTATAAAACGATTTTCAGGCAGATGATGAACGTTATACCGCAC 896
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
184 ePheSerIleIleArgAspLeuGlyAlaLeuAlaGlnValHisAlaGluA 201
: ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
897 CTAAGTGGCTCCCAAGAACCTGGCGGCTTGTCTATGTTTCATGCGGAA 946
||||| ||||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
201 snGlyAspIleValGluGluGlnLysArgLeuLeuGluGlyIle 217
||||| ||||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
947 ATGGGATGCTGATTGATTACTTACGAAAGGCGCTTCGGAAGGGAAT 996
||||| ||||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
218 ThrGlyProGluGlyHisValLeuSerHisProGluValGluAlaG 234
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
997 ACGAGCGCGATTTACCATGCTTTAACGCGCTCCAGAGTAGAAGGAGA 1046
||||| ||||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
234 uAlaValTyrArgAlaValThrIleAlaLysGlnAlaAsnCysProLeuT 251
||||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
1047 AGCGACGGCGCGCTGTCAATGTAGCAGAGCTTGGCGGTTTCACAACTT 1096
||||| ||||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
251 yrValThrLysValMetSerLysGlyAlaAlaAspAlaIleAlaGlnAla 267
||||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
1097 ACGTTGTACGTGACATGTGCGCAAGCGGTGGAAAAAATTCACACAGCG 1146
||||| ||||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
268 LysArgArgGlyValValPheGlyGluProIleThrAlaSerLeuG 284
||||| ||||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
1147 CGCAATAAAGGTGTGATGTGGGAGAACCGTGTCCGCAATATCTTGT 1196
||||| ||||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
284 yThrAspGlySerHisTyrTrpSerLysAsnTrpAlaLysAlaAlaAla 301
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1197 TCTCGACCAATCG...TATTTAGAAAAGCTGATTTTGAAGGCGCGAAT 1243
||||| ||||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
301 heValThrSerProValAsnProAspProThrThrAlaAspHisLeu 317
||||| ||||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
1244 ATGTTGGTCCCTCCGCTCGT...GAAATAATGGCATCAAGAGTATTG 1290
||||| ||||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
318 ThrCysLeuLeuSerSerGlyAspLeuGlnValThrGlySerAlaHisC 334
||||| ||||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
1291 TGGAAATCGCTGAAAAACGCGCAGCTGCAACCGCTTGGATCGGACCAATG 1340
||||| ||||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
334 sThrPheThr...ThrAlaGlnLysAlaValGlyLysAspAsnPheAla 350
||||| ||||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
1341 TTCATTGTGACTTTAAAGGCCAAAAAGAACTTGCAGAGGAGATTTACTA 1390
||||| ||||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
350 euIleProGluGlyThrAnGlyIleGluGluArgMetSerMetValTrp 366
||||| ||||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
1391 AAATTCCAACGGCGCGCGATGTCGAGGATCGGGTCAGCATTTCTTTC 1440
||||| ||||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
367 GluLysCysValAlaSerGlyLysMetAspGluAsnGluPheValAlaVa 383
||||| ||||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
1441 AGTGAAGGGGTAAAAAAGGAAGATACGTTAATCAATTTGTTCGATAT 1490
||||| ||||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
383 lThrSerThrAsnAlaAlaLysIlePheAsnPheTyrProArgLysGlyA 400
||||| ||||| : ||| : ||| : ||| : ||| : ||| : ||| : |||

```

```

: ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : |||||
1491 TATGTCGACAAGAAATTCGCAAAATTTGTCGGTTATTCCCGAGAAAGGAA 1540
400 rgValAlaValGlySerAspAlaAspLeuValIleTrpAsnProLysAla 416
: ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : |||||
1541 CGATCGCGTAGTTTCAGACGAGACTTAGTCATTTTTCACCGCGATATC 1590
417 ThrLysIleSerAlaLysThrHisAsnLeuAsnValGluTyrAsnI 433
: ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : |||||
1591 GAACGGGTGATTTCGCGGAAACACACCATATGCGCGTCGACTATATGTC 1640
433 ePheGluGlyValGluCysArgGlyAlaProAlaValValIleSerGln 450
||||| ||||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
1641 ATTGAAGGAATGAAGTAACGGGTCAACCGGTATCGGTTCG.TGCAGAG 1689
450 lYArgValAlaLeuGluAspGlyLysMetPheValThrProGlyAlaGly 466
||||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
1690 GCGAATTTGTTGTCGCTGATAACAATTTGTCGGAAACCGGGTACGGC 1739
467 ArgPheValProArgLysThrPhe 474
: ||||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
1740 CAATATTTAAACCGGCAAAATAC 1763

```

seq_name: /SIDS8/gcgdata/geneseq/geneseq/NA1999.DAT:AA13316

seq_documentation_block:

ID AAX13316 standard; DNA; 13085 BP.

XX AAX13316;

XX 19-MAR-1999 (first entry)

XX Enterococcus faecalis genome contig SEQ ID NO:379.

XX Enterococcus faecalis; contig; detection; Enterococcal infection;

XX vaccine; attenuation; computer readable medium; ds.

XX Enterococcus faecalis.

XX WO9850555-A2.

XX 12-NOV-1998.

XX 04-MAY-1998; 98WO-US08985.

XX 14-NOV-1997; 97US-0066009.

XX 06-MAY-1997; 97US-0044031.

XX 16-MAY-1997; 97US-0046655.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Barash SC, Dillon PJ, Kunsch CA;

XX WPI; 1999-045171/04.

XX New isolated Enterococcus faecalis polynucleotides and polypeptides
- used to develop products for the detection of Enterococcus and for
use in vaccines for prevention or attenuation of Enterococcus
infection.

XX Claim 1; Page 1549-1556; 2084pp; English.

XX A computer readable medium has been developed which has recorded on it
982 nucleotide sequences isolated from the Enterococcus faecalis genome.
XX AAX12938 to AAX13919 represent these nucleotide sequences which are
primary nucleotide sequences, also known as contigs. The computer-based
system can identify fragments of the Enterococcus faecalis genome with
commercial importance. The products can be used to detect the presence
of Enterococcus faecalis in samples. They can also be used for
diagnosing Enterococcal infection in an animal and monitoring
progression of disease, and for identifying agents which can be used to
modulate the growth or pathogenicity of Enterococcus faecalis, or
another related organism, in vivo or in vitro. In particular the

Tue Jul 31 13:08:18 2001

CC polypeptides encoded by the Enterococcus faecalis nucleotide sequences
 CC can be used in vaccines to prevent or attenuate an Enterococcal
 CC infection.
 XX
 SQ Sequence 13085 BP; 3789 A; 2961 C; 2106 G; 4222 T; 7 other;

alignment_scores:
 Quality 698.50 Length: 462
 Ratio: 2.268 Gaps: 9
 Percent Similarity: 66.667 Percent Identity: 34.199

alignment_block:
 US-09-367-496-8 x AAX13316/rev ..

Align seg 1/1 to reverse of: AAX13316 from: 1 to: 13085

17 LeuLeuileAArgGlyGlyArgileValAlaAspGlnSerPheTyrAl 33
 :::::::::::::::::::: :::::::::::::: ::
 8239 ATTCTGCTCAAGGTGGAACCGTCTGATCATATAATCGTCAATT 8190
 33 aaspValHisValGluAspGlyLeuileLysGlnIleGlyGluAsnLeuI 50
 :::::::::::::::::::: :::::::::::::: ::
 8189 GCATATTCGAATTACGGCGAGCAAAATTGTAGAGATGGGGCAAAATTAC 8140
 5C leValProGlyIleHisThrIleAspAlaHisGlyLeuMetValLeu 66
 :::::::::::::::::::: :::::::::::::: ::
 8139 CAGTAGAAATTTCT...CAGATTGAGGATGTACGGGTGTATGATTA 8093
 67 ProGlyGlyValAspValHisThrArgLeuGlnMet.....ProValIle 81
 :::::::::::::::::::: :::::::::::::: ::
 8092 CCAGGCTTTATGATCGCATACGATCTGGAATTAATAATGGAAGG 8043
 81 uGlyMetThrProAlaAspPheCysGlnGlyThrLysAlaLeuAla 98
 :::::::::::::::::::: :::::::::::::: ::
 8042 TTCGTTAAGTACGGCGGATTAATTTACCTGGAAGCAAGCGCGTGG 7993
 98 lagGlyThrThrMetIleLeuAspHisValPheProaspThrGlyVal 114
 :::::::::::::::::::: :::::::::::::: ::
 7992 CCAAGGTAGCTACGTTATGATGATGCTACACCTAATAAAGCGGC 7943
 115 SerLeuLeuAlaAlaThrGluGlnTTPArgGluArgAlaAspSerAla 131
 :::::::::::::::::::: :::::::::::::: ::
 7942 TCATTTGAAGACGTGTTAGTACTTGGAAATCAATGCTGGAAGCAAG 7893
 131 aCysCysAspTyrSerLeuHisValAspIleThrArgTTPHisGluSerI 148
 :::::::::::::::::::: :::::::::::::: ::
 7892 TTCATGTGATTTACTTACCATATGTCATGATGTAATGAAACCAACGA 7843
 148 leLysGluGluLeuAlaLeuValLysGlyLysGluLysGluLysSerPhe 164
 :::::::::::::::::::: :::::::::::::: ::
 7842 TAGCTGTGAATTCAGGAATGATT...GCTGAGGATATTCCTCTTC 7796
 165 LeuValPheMetAlaTyrLysAspArgCysGlnCysSerAspSerGlnMe 181
 :::::::::::::::::::: :::::::::::::: ::
 7795 AAATGTATATGGCTTAT...GATAATTGGGACCAACAGATGCGGAAT 7749
 181 tTyrGluIlePheSerIleLeuArgAspLeuGlyAlaLeuAlaGlnValH 198
 :::::::::::::::::::: :::::::::::::: ::
 7748 TTTTGAAGCGATGAAGGAATAAAGATTAATGGAATGTTAGCGGTTTC 7699
 198 IsAlaGluAsnGlyAspIleValGluGluGlnLysArgLeuLeuGlu 214
 :::::::::::::::::::: :::::::::::::: ::
 7698 APTGTGAATATGGTGTATTTAGTAGATGAATTAATCAATCATATGTTCT 7649
 215 LeuGlyIleThrGlyProGlyHisValLeuSerHisProGluGluVal 231
 :::::::::::::::::::: :::::::::::::: ::
 7648 CAAGGAAGCTAACCCGCGCATTCATCCGTTATCAAGACCGCGGCTGT 7599
 231 lGluAlaGluAlaValTyrArgAlaValThrIleAlaLysGlnAlaAsnC 248
 :::::::::::::::::::: :::::::::::::: ::
 7598 TGAAGCAGAAAGCGGTAGCAAGATATTTGATGATGCGGAATGCTGATT 7549

248 ysProLeuTyrValThrLysValMetSerLysGlyAlaAlaAspAlaIle 264
 :::::::::::::::::::: :::::::::::::: ::
 7548 TGCCGCTTAATATTGTCATTTAAGCACAACAGATCATTAAGACGGTG 7499
 265 AlaGlnAlaLysArgArgGlyValValPheGlyGluProIleThrAl 281
 :::::::::::::::::::: :::::::::::::: ::
 7498 GAACGGGCTCGTCAAGAGGACCAATCAGTCTATGTAGAAACTGCGCTCA 7449
 281 aserLeuGlyThrAspGlySerHisTyrTrpSerLysAsnTrpAlaLysA 298
 :::::::::::::::::::: :::::::::::::: ::
 7448 ATATCTTTTACTAGATCATCTGTATGATGCGCTAATTTT...GAAG 7402
 298 laAlaAlaPheValThrSerProValAsnProAspProThrThrAla 314
 :::::::::::::::::::: :::::::::::::: ::
 7401 CAGCAAAATATGCTCTCACCGCCACTACGC.....AAACGA 7364
 315 asHisLeuThrCysLeu.....LeuSerSerGlyAspLeuGlnVa 328
 :::::::::::::::::::: :::::::::::::: ::
 7363 GAGGATCAACGGGCTTTGTGCAAGGATTAAGAGGGGCAATCAATAC 7314
 328 lThrGlySerAlaHisCysThrPheThr...ThrAlaGlnLysAlaValG 344
 :::::::::::::::::::: :::::::::::::: ::
 7313 AATATCAGCAGATCATCTGCTGTTTAAATTTTACGGACAAAACACTTG 7264
 344 lLysAspAsnPheAlaLeuIleProGluGlyThrAsnGlyIleGluGlu 360
 :::::::::::::::::::: :::::::::::::: ::
 7263 CCAAGATGATTTAGTAAATTTCCCAATGGATGCCAGGTGTTCAACG 7214
 361 ArgMetSerMetValTrpGluLysCysValAlaSerGlyLysMetAspG 377
 :::::::::::::::::::: :::::::::::::: ::
 7213 AGACGAGAATTAATTTACACGGAAGGTGCGGAAGGCGCAATCACCTT 7164
 377 uAsnGluPheValAlaValThrSerThrAsnAlaAlaLysIlePheAsn 394
 :::::::::::::::::::: :::::::::::::: ::
 7163 GGAATAAATGTTGCTGCTTCTGAAATATTCCTAAACAATTCGGA 7114
 394 heTyrProArgLysGlyArgValAlaValGlySerAspAlaLeuVal 410
 :::::::::::::::::::: :::::::::::::: ::
 7113 TGTATCCACAAAAGGTGTGTGCAAGAGCGAGTGTGCGCATTTAGTT 7064
 411 lIleTyrAsnProLysAlaThrLysIleIleSerAlaLysThrHisAsn 427
 :::::::::::::::::::: :::::::::::::: ::
 7063 GTTGGGATCCCAACCAACCGCGGTGATGCTGCGAGAAGCAGTTACA 7014
 427 uAsnValGluTyrAsnIlePheGluGlyValGluCysArgGlyAlaPro 444
 :::::::::::::::::::: :::::::::::::: ::
 7013 AATATGCGATTTATACCGCTATGAGGATTTGAAACACCAAGGACAGGCTC 5964
 444 laValIleSerGlnGlyArgValAlaLeuGluAspGlyLysMetPhe 460
 :::::::::::::::::::: :::::::::::::: ::
 6963 GAATGCTTTTAAAGAGTCAAAAAGTTGCTCAAGCTGGCAAGTC... 6917
 461 ValThrProGlyAlaGlyArgPheValProArgLys 472
 :::::::::::::::::::: :::::::::::::: ::
 6916 ATCTTGCAAAATCAAGGAAATTTGTTTTCGTAAA 6881

seq_name: /SID8/gcgdata/geneseq/geneseq/NA1995.DAT:AA04198

seq_documentation_block:

ID AA04198 standard; DNA; 1374 BP.

XX

AC AA04198;

XX

DT 13-MAR-1996 (first entry)

XX

DE Agrobacterium radiobacter D-hydantoinase gene.

XX

KW D-hydantoinase; D-N-carbamylase; enzyme; stereospecific reaction;

KW D-amino acid; ss.

XX

OS Agrobacterium radiobacter.

XX

XX Location/Qualifiers

FK key

PN	WO9620275-A1.
XX	
PD	04-JUL-1996.
XX	
PF	26-DEC-1995; 95WO-JP02688.
XX	
PR	28-DEC-1994; 94JP-0326865.
XX	
PA	(KANF) KANEGAFUCHI KAGAKU KOGYO KK.
XX	
PI	Ikenaka Y, Nanba H, Takahashi S, Takano M, Yajima K;
PT	Yamada Y;
XX	
DR	WPI; 1996-321848/32.
XX	
PT	P-PSDB; AAW03024.
XX	
PT	Prodn. of D-N-carbamoyl amino acid from 5-substd. hydantoin - using
PT	a recombinant hydantoinase derived from a strain of Pseudomonas,
XX	Agrobacterium or Bacillus
XX	
PS	Claim 13; Page 30-35; 54pp; English.
XX	
CC	D-N-carbamoyl-alpha-amino acid is produced from a 5-substituted
CC	hydantoin by treatment with a hydantoinase expressed by a
CC	transformant microorganism carrying a vector containing DNA coding
CC	for the hydantoinase and derived from Bacillus sp. KNK245,
CC	Agrobacterium sp. KNK712 or Pseudomonas sp. KNK003A. The D-N-
CC	carbamoyl-alpha-amino acid can be used for the production of
CC	optically active alpha amino acids (especially D-phenylglycine and
CC	p-hydroxyphenylglycine) as intermediates for drug synthesis,
CC	especially for the production of semi-synthetic penicillin and
CC	cephalosporin antibiotics.
XX	
SQ	Sequence 3190 BP; 639 A; 890 C; 861 G; 536 T; 16 U; 248 other:

```

alignment_scores:
  Quality: 684.50      Length: 470
  Ratio: 2.244        Gaps: 9
  Percent Similarity: 64.894  Percent Identity: 34.255

alignment_block:
  US-09-367-496 x AAT31257  ..

  Align seg 1/1 to: AAT31257 from: 1 to: 3190

17  LeuLeuIleAArgGlyGlyArgIleValAlaSnAspGlnSerPheTyrAl 33
    ::::||||:::|||||  |||||:::  |||
166 ATCATTTATCAAAACGGAACCATCTGTACCGGGATGGCATTTCTCGCGC 215
    ::::||||:::|||||  |||||:::  |||

33  aspValHisValGluAspGlyLeuIleLysGlnIleGlyGluAsnLeuI 50
    :|||:::|||||  |||||:::  |||||
216 CGATCTCGGGATCAAGATGGCAAGATCAACCCAGATATCGGC..... 255

50  levalProGlyGlyIle.....HisThrIleAspAlaHisGly 62
    |||||:::|||||  |||||:::  |||||
256 .....GGCGCGTTCGGCCCGAGCGAGCGACGATCGACGCGCGCGC 297

63  LeuMetValLeuProGlyGlyValAspValHisThrArgLeuGlnMetPr 79
    |||||:::|||||  |||||:::  |||||
298 CGCTAGTGCTTTCGGGGCGGATACAGCTTCACAGCATGTCTGAAACCGT 347

79  oValLeuGlyMetThrProAlaAspPheCysGlnGlyThrLysAlaA 96
    ::::||||  |||||:::  |||||
348 CAGCTTCAACAGCAGTCGCGCGGACACGTCCTGCAACACGCGCGTTCGGC 397

96  laLeuAlaGlyGlyThrThrMetIleLeuAspHisValPheProAspThr 112
    |||||:::|||||  |||||:::  |||||
398 CGCGCTGNGCGGAACGACACCATCTGCTGATTTCTGTACACGAGATCGC 447

113 GlyValSerLeuLeuAlaAlaTyrGluGlnTrpArgGluArgAlaAspSe 129
    :|||:::|||||  |||||:::  |||||
448 GGCCACACCCCTGGCGGAACCGCTCCCAAGTGGACGATGTGCGCGCGG 497
    :|||:::|||||  |||||:::  |||||

```

```
seq_name: /STD88/gcgdata/geneseq/NA1996.DAT:AAT31257
seq_documentation_block:
ID      AAT31257 standard; DNA; 3190 BP.
XX
XX      AAT31257;
XX
XX      31-JAN-1997 (first entry)
XX
XX      Hydantoinase coding sequence.
XX
XX      Hydantoinase; hydantoin; d-N-carbamoyl-alpha-amino acid;
XX      intermediate; drug synthesis; penicillin; cephalosporin;
XX      antibiotic; ss.
XX
XX      Agrobacterium sp.
XX
XX      Key      Location/Qualifiers
XX      CDS      160..1534
XX              /*tag= a
XX              /product= Hydantoinase
XX
```

421 rAlauLysThrHisAsnLeuValGlnTyrRsrNilePheGluGlyValG 438
 :
 1365 ACAGACCGCCATGCACAACCCTCATGGATTACTCCTCAAGGAGCACACA 1414
 :
 438 luCySArgGlyAlaProAlaValIleSerGlnGlyArgValAlaLeu 454
 :
 1415 AGGTCAAAGCGTGTCGAAGACGGTCTCCTCGTGGCAAGTTATCGTC 1464
 :
 455 GluaSPgLyMetPheValThrProGlyAlaGlyArgPheValProAr 471
 :
 1465 GACGAGGTTCCTATGTCGCCGAAACCGACGACGGAGAATAAATCCTGAACG 1514
 :
 471 gLYSThrPHe 474
 | :
 1515 TCGCAATAAC 1524
 :

seq_name: /SDSB/gcgdata/geneseq/geneseqn/NA2000.DAT.AAF21851

seq_documentation_block:
ID AAF21851 standard; DNA; 3039 BP.
XX
AC AAF21851;
XX
DT 27-MAR-2001 (first entry)
XX
DE Human breast and ovarian cancer associated antigen gene SEQ ID 238.

Human; breast cancer; ovarian cancer; cytostatic; immunosuppressive;
neurotic; neuroprotective; antiviral; antiallergic; hepatotropic;
KW antidabetic; antiinflammatory; anticancer; vulnarary; anticonvulsant;
KW antibacterial; antifungal; antiparasitic; cardiant; immune disorder;
KW Addison's disease; allergy; autoimmune haemolytic anaemia;
KW autoimmune thyroiditis; diabetes mellitus; Crohn's disease;
KW multiple sclerosis; rheumatoid arthritis; ulcerative colitis;
KW cardiovascular disorder; wound healing; neurological disease; ds.
XX Homo sapiens.
XX WO2000055173-A1.
PN XX
PD 21-SEP-2000.
XX
PF 08-MAR-2000; 2000WO-US05881.
XX PR
PR 12-MAR-1999; 99US-0124270.
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PI Rosen CA, Ruben SM;
XX WPI; 2000-611515/58.
DR P-PADB; AAB58948.
XX
PT New human breast and ovarian cancer associated gene sequences and the polypeptides encoded by these genes, useful in the prevention,
PT treatment and diagnosis of cancer, immune disorders, cardiovascular
PT disorders and neurological diseases -
XX Claim 1; Page 662-663; 1299pp; English.
PS Sequences AAF21614 - AAF22031 represent DNA sequences encoding human proteins AAB58711 - AAB59128. The DNA and protein sequences are CC associated with breast and ovarian cancer. Included in the invention are CC sequences AAF22032 - AAF22040 and AAB59129 which are used in the CC isolation and characterisation of the DNA and protein sequences of the CC invention. The breast and ovarian cancer associated DNA, protein, agonist or antagonist sequences exhibit cytotoxicity; immunosuppressive; CC neurotropic; neuroprotective; antiviral; antiallergic; hepatotropic; CC antidabetic; antiinflammatory; anticancer; vulnarary; anticonvulsant; CC antibacterial; antifungal; antiparasitic and cardiac activity. The CC polynucleotide and protein sequences are used in the diagnosis of cancer, particularly breast and ovarian cancer. The nucleic acid sequences, CC proteins, agonists and antagonists may also be used in the diagnosis,

CC prevention and treatment of immune disorders e.g. Addison's disease,
CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
CC arthritis and ulcerative colitis; cardiovascular disorders such as
CC myocardial ischaemias; wound healing; neurological diseases such as
CC cerebral anoxia and epilepsy; and infectious diseases.
XX
SQ Sequence 3039 BP; 744 A; 711 C; 659 G; 919 T; 6 other;

alignment_scores:

Quality: 600.00 Length: 166
Ratio: 4.225 Gaps: 0
Percent Similarity: 85.542 Percent Identity: 68.072

alignment_block:

US-09-367-496-8 x AAF21851 ..

Align seg 1/1 to: AAF21851 from: 1 to: 3039

406 AspAlaAspLeuValIleTrpAsnProLysAlaThrLysIleIleSerAl 422
|||||
3 GATGGCGACCTGGTCTATCTGGGACCCGACACGCGTTAAAGCATCTCTGC 52
422 aLysThrHisAsnLeuAsnValIleGluTyrAsnIlePheGluGlyValGluC 439
|||||
53 CAAGACACACACAGCTCTCTCGAGTACACATCTTTGAGGCGATGGAGT 102
439 ysArgGlyAlaProAlaValIleSerGlnGlyArgValAlaLeuGlu 455
|||||
103 GCCGGCGCTCCCCACTGGTGTGTCATCAGCCAGGGGAGATTGTCTGGAG 152
456 AspGlyLysMetPheValThrProGlyAlaGlyArgPheValProArgLy 472
|||||
153 GACGGCACCTGCATGTNACCGAANGCTCTGGACGCTACATTCCTCCCGAA 202
472 sThrPheProAspPheValTyrLysArgIleLysAlaArgAsnArgLeuA 489
|||||
203 GCCCTTCCTGATTTTATTACAGCGGTATCAGGCAAGGAGCAGGCTGG 252
489 laGluIleHisGlyValProArgGlyLeuTyrAspGlyProValHisGlu 505
|||||
253 CTGAGCTGAGAGGGGTCTCTCGTGGCTGTATCAGGAGCCCGTGTGTGAA 302
506 ValMetValProAlaLysProGlySerGlyAlaProAlaArgAlaSerCy 522
|||||
303 GTGTCTGTGAGGCCCAAGACAGTCACTCCAGCCTCTCTCGGCCAAGACGTC 352
522 sProGlyLysIleSerValProProValArgAsnLeuHisGlnSerGlyP 539
|||||
353 TCCTGCCAAGCAGCAGGCCCCACCTGTCCGGAACCTGCACCATCTGGAT 402
539 heSerLeuSerGlySerGlnAlaAspHisIleAlaArgThrAla 555
|||||
403 TCAGTTGT 500
556 GlnLysIleMetAlaProProGlyGlyArgSerAsnIleThrSerLeu 571
|||||
453 CAGCGTATCGTGGCGCCCGCCCGGTGGCGGTGCGCAACATCACACGCTG 500

28	92	3.1	1421	3	US-09-335-409-2	Sequence 2, Appl
29	92	3.1	2115	4	US-09-324-867-5	Sequence 5, Appl
30	91.5	3.1	2133	2	US-08-670-707A-37	Sequence 37, Appl
31	91.5	3.1	2133	4	US-09-037-601-37	Sequence 37, Appl
32	91.5	3.1	3170	4	US-09-036-987A-4	Sequence 4, Appl
33	90.5	3.0	652	1	US-08-318-831-8	Sequence 8, Appl
34	90.5	3.0	1780	1	US-08-769-309A-5	Sequence 5, Appl
35	90.5	3.0	1928	3	US-09-034-570-5	Sequence 5, Appl
36	89	3.0	4928	4	US-08-936-987A-5	Sequence 5, Appl
37	88.5	3.0	365	2	US-08-846-762-80	Sequence 80, Appl
38	88.5	3.0	365	2	US-08-576-626A-46	Sequence 46, Appl
39	88.5	3.0	397	1	US-08-220-448A-76	Sequence 76, Appl
40	88.5	3.0	397	1	US-08-230-448A-76	Sequence 76, Appl
41	88.5	3.0	397	1	US-08-175-069A-76	Sequence 76, Appl
42	88.5	3.0	464	2	US-09-021-323-3	Sequence 3, Appl
43	88.5	3.0	948	1	US-08-698-551-14	Sequence 14, Appl
44	88.5	3.0	948	2	US-08-602-228-14	Sequence 14, Appl
45	88.5	3.0	948	2	US-08-839-032A-14	Sequence 14, Appl

ALIGNMENTS

```

1
RESULT
US-08-289-703-1
; Sequence 1, Application US/08289709
; Patent No. 5523224
;
; GENERAL INFORMATION:
;
; APPLICANT: Burtischer, Helmut; Lang, Gunter; Popp, Friedrich
;
; TITLE OF INVENTION: Recombinant D-Hydantoinase, A Process For the
;
; TITLE OF INVENTION: Production and Use
;
; NUMBER OF SEQUENCES: 4
;
; CORRESPONDENCE ADDRESS:
;
; ADDRESSEE: Felfe & Lynch
;
; STREET: 805 Third Avenue
;
; CITY: New York City
;
; STATE: New York
;
; COUNTRY: USA
;
; ZIP: 10022
;
; COMPUTER READABLE FORM:
;
; MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage

```

SUMMARIES

Result No.	Query		DB	ID	Description
	Score	Match Length			
1	826	27.7	460	1	US-08-289-709-1
2	826	27.7	460	1	US-08-602-656-1
3	196	6.6	568	1	US-07-732-242C-3
4	142	4.8	567	1	US-08-967-513-5
5	142	4.8	567	2	US-08-687-645B-5
6	136.5	4.6	566	2	US-08-920-095-3
7	136.5	4.6	566	5	PCT-US96-US0800-3
8	136.5	4.6	569	2	US-08-467-832-21
9	136.5	4.6	569	4	US-08-433-697-21
10	134.5	4.5	569	2	US-08-467-832-26
11	134.5	4.5	569	4	US-08-432-697-26
12	129	4.3	840	2	US-08-467-822-25
13	129	4.3	840	4	US-08-432-697-25
14	112.5	3.8	2616	6	5206163-3
15	102.5	3.4	1811	2	US-08-804-273C-5
16	102	3.4	3567	2	US-07-642-724C-4
17	102	3.4	3567	3	US-08-439-009A-4
18	99	3.3	569	2	US-08-467-822-27
19	99	3.3	569	4	US-08-432-697-27
20	99	3.3	7257	3	US-09-335-409-5
21	97.5	3.3	841	1	US-08-350-884-86
22	97.5	3.3	841	1	US-08-709-173-86
23	97.5	3.3	841	2	US-08-709-177-86
24	94	3.1	3739	3	US-09-320-878-2
25	93	3.1	4545	2	US-08-804-227C-14
26	93	3.1	4550	2	US-08-804-227C-8
27	93	3.1	4550	2	US-08-804-198-2
					Sequence 1, Appl
					Sequence 1, Appl
					Sequence 3, Appl
					Sequence 5, Appl
					Sequence 3, Appl
					Sequence 21, Appl
					Sequence 21, Appl
					Sequence 26, Appl
					Sequence 26, Appl
					Sequence 25, Appl
					Sequence 25, Appl
					Patent No. 5206163
					Sequence 5, Appl
					Sequence 4, Appl
					Sequence 4, Appl
					Sequence 27, Appl
					Sequence 27, Appl
					Sequence 5, Appl
					Sequence 86, Appl
					Sequence 86, Appl
					Sequence 86, Appl
					Sequence 2, Appl
					Sequence 14, Appl
					Sequence 8, Appl
					Sequence 2, Appl

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2001, 11:36:16 ; Search time 12.3 Seconds
(without alignments)
957.535 Million cell updates/sec

US-09-367-496-8
2985
Title: RAKOIMAPPGRSNTISLS 572
perfect score: 1 WROCKKVSIBETTSDBIITP

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 197339 seqs, 20590346 residues

Total number of hits satisfying chosen parameters: 197339

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Minimum Match 08

Post-processing: Minimum Match 100%
Maximum Match 100%
Listing first 45 summaries

```
Database : Issued_Patents_AA.*
1: /cgn2_6/ptodata/1/iaa/5A_COMB pep: *
2: /cgn2_6/ptodata/1/iaa/5B_COMB pep: *
3: /cgn2_6/ptodata/1/iaa/6A_COMB pep: *
4: /cgn2_6/ptodata/1/iaa/6B_COMB pep: *
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB pep: *
6: /cgn2_6/ptodata/1/iaa/backfiles1 pep: *
```

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Query Match 27.7%; Score 826; DB 1; Length 460;
Best Local Similarity 39.7%; Pred. No. 2.6e-78;
Matches 182; Conservative 194; Indels 4; Gaps 4;
18 LIRGRIVNDQSFVADYHVEDGLIKQIGENILVPGGIHTIDAHGLMVLPGGVVHTRLQ 77

```

Db 4 IIKNGTIVTADTYEADLLIKDKIAMIGQHL-EKGAEVIDAKGCVYFPGGIDSHTHLD 62
QY 78 MPVLGMPADDFCQCTKAALAGGTTMILDHVPDPDGVSLAAYEQRERADSAACCDYSL 137
Db 63 MPFGTGTVDKDFESCTIAAAGGTTTIDFCLTNKGEPLKKAIEIWHNKAAGKAVIDYGF 122
QY 138 HVDITRWHESKEELEALVKEGVNSFLVFMAYKDRCCSDSQMSYEIFSIIRDLGALAQV 197
Db 123 HLMISEITDDVLEELPKVIAEEGITSFKVFMAYKNVFOADDGTLRTLVAAKELGALVMV 182
QY 198 HAENGDIIVEEQKRLLELGITGPEGHVLSHPVEAEAVYRAVITAKOANPLVYTKVMS 257
Db 183 HAENGDIVIDLTKKALAEAGNTEPIYHALTRPEVEGEATGRACQTELAGSOLYVHVTC 242
QY 258 KGAADATAQAKRRGVVFGPEPITASLGTDGSHYWSKNWAKAAAFVTPSPVNPDPDTADHL 317
Db 243 QAIVEKTAQARNKGLDVMGETCPOYLVDQS-YLEKPDFEGAKYVWSPPLR-EKWHQEV 300
QY 318 TLLSSDGLQVTSAGHCTFT-TAQKAVGKDNFALIPGTTNGIERMSMWKCVASGKMD 376
Db 301 WNALKNGQLQTLGSDQCSDFKQKELGRGDTKIPNGGPMVEDRVSLFSEGVKKGRIT 360
QY 377 ENFVAVTSTNAAKIFNFPYPRKGRVAVGSADLVVWPKATKIISAKTHNLNVEYNIFEG 436
Db 361 LNQFVDMSTRIAKLFLGFPKGTIAVGSADLVIFDPDIERSISAETHHMAVDYNAFEG 420
QY 437 VECRGAPAVVISQGRVALEDGKMFVTPGAGRFVPRKTF 474
Db 421 MKVTGEPVSLCRGEFVVRDKQFVGKPGQYLKAGCF 458

```

RESULT 2

```

US-08-602-656-1
; Sequence 1, Application US/08602656
; Patent No. 5679571
; GENERAL INFORMATION:
; APPLICANT: Burtischer, Helmut; Lang, Gunter; Popp, Friedrich
; TITLE OF INVENTION: Recombinant D-Hydantoinase, A Process For the
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/602,656
; FILING DATE: 16-FEB-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/289,709
; FILING DATE: 12-AUGUST-1994
; APPLICATION NUMBER: P 43 28 829.4
; FILING DATE: 27-AUGUST-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, No. 5679571man D.
; REGISTRATION NUMBER: 30,946
; REFERENCE/DOCKET NUMBER: BOER 1041
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 460 amino acids
; TYPE: amino acid
; STRANDEDNESS: single

```

```

; TOPOLOGY: linear
; US-08-602-656-1

```

```

Query Match 27.7%; Score 826; DB 1; Length 460;
Best Local Similarity 39.7%; Pred. NO. 2.6e-78;
Matches 182; Conservative 78; Mismatches 194; Indels 4; Gaps 4;

QY 18 LIRGRIYNDQDSFYADVHVEDGLIKQIGENLIVPGGTHITDAHGLMVLPGGVVHTRIQ 77
Db 4 IIKNGTIVTADTYEADLLIKDKIAMIGQHL-EKGAEVIDAKGCVYFPGGIDSHTHLD 62
QY 78 MPVLGMPADDFCQCTKAALAGGTTMILDHVPDPDGVSLAAYEQRERADSAACCDYSL 137
Db 63 MPFGTGTVDKDFESCTIAAAGGTTTIDFCLTNKGEPLKKAIEIWHNKAAGKAVIDYGF 122
QY 138 HVDITRWHESKEELEALVKEGVNSFLVFMAYKDRCCSDSQMSYEIFSIIRDLGALAQV 197
Db 123 HLMISEITDDVLEELPKVIAEEGITSFKVFMAYKNVFOADDGTLRTLVAAKELGALVMV 182
QY 198 HAENGDIIVEEQKRLLELGITGPEGHVLSHPVEAEAVYRAVITAKOANPLVYTKVMS 257
Db 183 HAENGDIVIDLTKKALAEAGNTEPIYHALTRPEVEGEATGRACQTELAGSOLYVHVTC 242
QY 258 KGAADATAQAKRRGVVFGPEPITASLGTDGSHYWSKNWAKAAAFVTPSPVNPDPDTADHL 317
Db 243 QAIVEKTAQARNKGLDVMGETCPOYLVDQS-YLEKPDFEGAKYVWSPPLR-EKWHQEV 300
QY 318 TLLSSDGLQVTSAGHCTFT-TAQKAVGKDNFALIPGTTNGIERMSMWKCVASGKMD 376
Db 301 WNALKNGQLQTLGSDQCSDFKQKELGRGDTKIPNGGPMVEDRVSLFSEGVKKGRIT 360
QY 377 ENFVAVTSTNAAKIFNFPYPRKGRVAVGSADLVVWPKATKIISAKTHNLNVEYNIFEG 436
Db 361 LNQFVDMSTRIAKLFLGFPKGTIAVGSADLVIFDPDIERSISAETHHMAVDYNAFEG 420
QY 437 VECRGAPAVVISQGRVALEDGKMFVTPGAGRFVPRKTF 474
Db 421 MKVTGEPVSLCRGEFVVRDKQFVGKPGQYLKAGCF 458

```

RESULT 3

```

US-07-732-242C-3
; Sequence 3, Application US/07732242C
; Patent No. 5298399
; GENERAL INFORMATION:
; APPLICANT: Uozumi, Takeshi; Masaki, Haruhiko;
; APPLICANT: Hidaka, Makoto; Nakamura, Akira;
; APPLICANT: Maeda, Michihisa; Yoneta, Yasuo
; TITLE OF INVENTION: Gene of Urease
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Frischauf, Holtz, Goodman & Woodward, P.C.
; STREET: 600 Third Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10016-2088
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 0.72mb
; COMPUTER: IBM PC compatible (NEC PC-9801 RX)
; OPERATING SYSTEM: MS DOS
; SOFTWARE: ASCII Form
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/732,242C
; FILING DATE: 19910718
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JPN 2-210178
; FILING DATE: 10-AUG-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Goodman, Herbert
; REGISTRATION NUMBER: 17081

```

REFERENCE/DOCKET NUMBER: 910532/HG
TELEPHONE: (212)972-1400
TELEFAX: (212)370-1622
TELEX: 236268
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 568 amino acids
TYPE: AMINO ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-07-732-242C-3

Query Match 6.6%; Score 196; DB 1; Length 568;
Best Local Similarity 19.8%; Pred. No. 1.2e-11;
Matches 122; Conservative 83; Mismatches 211; Indels 200; Gaps 26;
QY 1 MSFOGKSI-----PRITSDR---LLIRGRIVDDQSFYADVHVHVEDGLIKOIGE--- 47
Db 43 VFGGKVRIDGMCQHPLATSDCEVDLVLTNALIIVDTYGIKADIGIKDGMASIGKAGN 102
QY 48 -----NLIVPGGIHTIDAGHLMVPGGVVHTRLOMPVLGWTTPADDFCQGTAAALAGG 100
Db 103 PLLMDGVDMVIGAAETVIAAEGMIVTAGGIDAHIFICP-----QQIETALASG 151
QY 101 -TTIILHVFDPDTGVSLLAAVEQWRERADSAACDYSLVHVDITRWHSIKEELEALVKEK 159
Db 152 VTTMIGGTGTATGTC-----NATTC-----TPGPNHRLQA----- 184
QY 160 GVSFLFVMAVYKDRCCSDSQMYEIFSIIRDLGALAAQVHAENGDIIVEEQKRLLELGITG 219
Db 185 -AEEFNLGFLGKNCSD-----EAPLKEQIEAGAVG 216
QY 220 PEGHVLSPHEE--VBAEAVYRAVTIAKQANGLYV-TKVMKG--AADAIAQAKRRGVVV 274
Db 217 LKLH-----EDWGSTAAIDTCLKVADRYDVOVAIHTDLNEGGVEVETLKAIDGRVHT 271
QY 275 FCEPITASLGTDGSHWKNWAKAAAFVTSPPVNDP-----TTADHLTCLLSSGDLQ 327
Db 272 Y-----HTEGAGGH--APDIKAAGFPNLPSTNTPRYTINTLEHDLMLVCHHLD 324
QY 328 VTGSAHCTTTAAQAVKND-----FALIPETNG---IEERMSVWKEC--- 369
Db 325 ANIPEDIAF--ADSRIRKETIAAEDVLHDLGVFSMISSDSQAMGRVGEVITITWTADKM 382
QY 370 -VASGKMDEN-----EFVAVTSTNAKIFNFYPRKGRVAVGSDADLVINPDKATK 418
Db 383 KQORGKLEDNGVDGNFRVKRYIAKYNPAIAGHADIYGVSEVGKLDLVVWNP--- 438
QY 419 IISAKTHNLNVEYNIEFEGVECRGAPAVVISQGRVALED-----CKMFTV 462
Db 439 -----AFFGVK-----PELVKGGMIATMTGDPNASIPTPPVLYRPMFAA 480
QY 463 PGAGR-----FVPRFTFDPFVYKRIKARNLAEIHGV-----PRGLYDGP 502
Db 481 KGDARYQTSTTFVSKAAAYEGKIEHQGLKKKVPVHGIRKLYKKDLILNDKTPKIDVDQ 540
QY 503 VHEVMVPKPGSGAPA 518
Db 541 TYEVKVDGQLVTCPEA 556

RESULT 4
US-08-967-513-5
Sequence 5, Application US/08967513
Patent No. 5783436
GENERAL INFORMATION:
APPLICANT: Robert P. Hausinger
TITLE OF INVENTION: Mutant Urease and Method
TITLE OF INVENTION: of Use For Determination
TITLE OF INVENTION: of Urea

NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Ian C. McLeod
STREET: 2190 Commons Parkway
CITY: Okemos
STATE: Michigan
COUNTRY: USA
ZIP: 48864
COMPUTER READABLE FORM:
MEDIUM TYPE: diskette, 5.25 inch, 360 KB
MEDIUM TYPE: storage
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/967,513
FILING DATE: 11-NOV-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/687,645
FILING DATE: July 26, 1996
ATTORNEY/AGENT INFORMATION:
NAME: Ian C. McLeod
REGISTRATION NUMBER: 20,931
REFERENCE/DOCKET NUMBER: MSU 4.1-309
TELECOMMUNICATION INFORMATION:
TELEPHONE: (517) 347-4100
TELEFAX: (517) 347-4103
TELEX: NO. 5783436
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 567
TYPE: amino acids
STRANDEDNESS: Single
TOPOLOGY: Linear
MOLECULE TYPE:
DESCRIPTION: protein
HYPOTHETICAL: No
ANTI-SENSE: No
ORIGINAL SOURCE:
ORGANISM: N/A
STRAIN: N/A
INDIVIDUAL ISOLATE: N/A
CELL TYPE: N/A
FEATURE:
NAME/KEY: subunit UreC
LOCATION:
IDENTIFICATION METHOD: Sequencing
OTHER INFORMATION: Encoded subunit of mutant
OTHER INFORMATION: urease
US-08-967-513-5

Query Match 4.8%; Score 142; DB 1; Length 567;
Best Local Similarity 20.8%; Pred. No. 5.9e-06;
Matches 122; Conservative 69; Mismatches 212; Indels 184; Gaps 24;
QY 17 LLIRGRIVDDQSFYADVHVHVEDGLIKOIGE-----NLIVPGGIHT--IDAHLAVL 66
Db 68 LVLTNALIVDHGIVKADIGVKGDRIFAGKAGNPDIOFNPVTIPIGAATEVIAAEGKIYT 127
QY 67 PGVDVHTRLOMPVLGWTTPADDFCQGTAAALAGGTTMILDHVPDPTGVSLLAAYEQWRER 126
Db 128 AGGIDTHIHWCIPQ-----QAEALVSGVTTWGGGTGPAAGT----- 165
QY 127 ADSAACCDYSLHVDITRWHSIKEELEALVKEGVNSFLVFMAYKDRCCSDSQMYEIFS 186
Db 166 -HATTCFPG-----PWYIS-----RMLQAAD-----S 186
QY 187 IIRDLGALAAQVHAENGDIIVEEQKRLLELGITGPEGHVLSPHEVEA--EAVYRAVTIAK 244
Db 187 LPVNIIGLKGNSQPDALREQ-----VAAGVIG-----LKIQDWGATPAADICALTVAD 237

```

QY 245 QANCPYV---TKVMSKGAADAIAQARRGVVVFGEPIITASLGTDGSHVSKNWKAAAF 301
Db 238 EMDIQVALHSDTLNESGFVEDTLAAIGRTIHTF-----HTEGAGGHH--APDIITACAH 290
QY 302 VTSPPVNPDPPTA-----DHLTCLSSDGLQVTSAGHCTF-----TTAQKAVGKD 346
Db 291 PHILPSTNPTLPYTLNTIDEHLDMLVCHHLDPPDIAEDVAFASRRIRRETIAAEDVLHD 350
QY 347 --NFALIPEGTNG---IEERMSVMWEKC---VASGKMDK-----NEFVAVTSTN 387
Db 351 LGAFSLTSSDQAMGRVGEVILRTVOVAHRMKVQVQALAEETGDNDFRVKRYIAKYTN 410
QY 388 AAKIFNFPKRGVAVGSDADLVINPKATKIISAKTHNLNVEYNIFEGVCEGAPAVVI 447
Db 411 PALTHGIAHEVGSIEVGLADLVVWSP-----AFFGVK----PATVI 448
QY 448 SQGRVALED-----GKMFVTPGAGRFVPRKTF-----PDFVYKRIKAR 485
Db 449 KGMIAIAPMGDINASITPQPVHYRPMFGALGSARHCHRLTFLSQAAAANGVAERLNL 508
QY 486 NLAIEIHGV-----PRGLYDGPVHEVMVPAKPGSGAPA 518
Db 509 SAIAVVKGCRVTQKADVMHNSLQPNITVDAQTYEVRVDGELITSEPA 555

```

RESULT 5

US-08-687-645B-5

; Sequence 5, Application US/08687645B

; Patent No. 5846752

; GENERAL INFORMATION:

; APPLICANT: Robert P. Hausinger

; TITLE OF INVENTION: Mutant Urease and Method

; TITLE OF INVENTION: of Use For Determination

; TITLE OF INVENTION: of Urea

; NUMBER OF SEQUENCES: 5

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Ian C. McLeod

; STREET: 2190 Commons Parkway

; CITY: Okemos

; STATE: Michigan

; COUNTRY: USA

; ZIP: 48864

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette, 5.25 inch, 360 Kb

; MEDIUM TYPE: storage

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: MS-DOS

; SOFTWARE: Wordperfect 5.1

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/687.645B

; FILING DATE: July 26, 1996

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER:

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Ian C. McLeod

; REGISTRATION NUMBER: 20,931

; REFERENCE/DOCKET NUMBER: MSU 4.1-309

; TELEPHONE: (517) 347-4100

; TELEFAX: (517) 347-4103

; TELEX: No. 5846752e

; INFORMATION FOR SEQ ID NO: 5:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 567

; TYPE: amino acids

; STRANDEDNESS: Single

; TOPOLOGY: Linear

; MOLECULE TYPE:

; DESCRIPTION: protein

; HYPOTHETICAL: No

; ANTI-SENSE: No

```

; ORIGINAL SOURCE:
; ORGANISM: N/A
; STRAIN: N/A
; INDIVIDUAL ISOLATE: N/A
; CELL TYPE: N/A
; FEATURE:
; NAME/KEY: subunit UreC
; LOCATION:
; IDENTIFICATION METHOD: Sequencing
; OTHER INFORMATION: Encoded subunit of mutant
; OTHER INFORMATION: urease
; US-08-687-645B-5

```

Query Match

4.8%; Score 142; DB 2; Length 567;

Best Local Similarity 20.8%; Pred. No. 5.9e-06;

Matches 122; Conservative 69; Mismatches 212; Indels 184; Gaps 24;

```

QY 17 LLIRGGRIVDDQSFYADVHVHVEDGLIKOIGE-----NLIVPGGIHT--IDAHGLMLV 66
Db 68 LVTNALIVDHGIVKADIGVKDGRIFAIGKAGNPDIQPNVTIPIGAATEVIAEGKIYT 127
QY 67 PGVDVHTRLQMPVLGMPADDFCQGTKAALAGTTHLDHVFDDTGVSLLAAYEQWRER 126
Db 128 AGGIDTHIHICPQ-----QAEELVSGVTTVMVGGGTGPAAGT-----165
QY 127 ADSAACCDYSLHVDITRWHSIKEELEALVKEGVNSFLVMAYKDRCCQSDSQMYEIFS 186
Db 166 --HATTCTPG-----PWYIS-----RMLQAAD-----S 186
QY 187 IIRDLGALQVHAENGDIVEEQKRLLLELITGPEGHVLSHPEVEA--EAVYRAVTIAK 244
Db 187 LPVNIIGLLKGNVSQPDALREQ---VAAGVIG---LKIQEDWGATPAIDCALTVAD 237
QY 245 QANCPYV---TKVMSKGAADAIAQARRGVVVFGEPIITASLGTDGSHVSKNWKAAAF 301
Db 238 EMDIQVALHSDTLNESGFVEDTLAAIGRTIHTF-----HTEGAGGHH--APDIITACAH 290
QY 302 VTSPPVNPDPPTA-----DHLTCLSSDGLQVTSAGHCTF-----TTAQKAVGKD 346
Db 291 PHILPSTNPTLPYTLNTIDEHLDMLVCHHLDPPDIAEDVAFASRRIRRETIAAEDVLHD 350
QY 347 --NFALIPEGTNG---IEERMSVMWEKC---VASGKMDK-----NEFVAVTSTN 387
Db 351 LGAFSLTSSDQAMGRVGEVILRTVOVAHRMKVQVQALAEETGDNDFRVKRYIAKYTN 410
QY 388 AAKIFNFPKRGVAVGSDADLVINPKATKIISAKTHNLNVEYNIFEGVCEGAPAVVI 447
Db 411 PALTHGIAHEVGSIEVGLADLVVWSP-----AFFGVK----PATVI 448
QY 448 SQGRVALED-----GKMFVTPGAGRFVPRKTF-----PDFVYKRIKAR 485
Db 449 KGMIAIAPMGDINASITPQPVHYRPMFGALGSARHCHRLTFLSQAAAANGVAERLNL 508
QY 486 NLAIEIHGV-----PRGLYDGPVHEVMVPAKPGSGAPA 518
Db 509 SAIAVVKGCRVTQKADVMHNSLQPNITVDAQTYEVRVDGELITSEPA 555

```

RESULT 6

US-08-920-095-3

; Sequence 3, Application US/08920095

; Patent No. 5837240

; GENERAL INFORMATION:

; APPLICANT: Cynthia K. Lee et al.

; TITLE OF INVENTION: MULTIMERIC, RECOMBINANT UREASE VACCINE

; NUMBER OF SEQUENCES: 3

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Fish & Richardson P.C.

; STREET: 225 Franklin Street

; CITY: Boston

; STATE: MA

; COUNTRY: USA

Tue Jul 31 13:08:17 2001

546 ADDEIART-----AQIMAPPGRSNTSL 571
534 AHIEVNPETYHVVDGKEVTSKPAKVSLAQL 565

RESULT 7
PCT-US96-05800-3
Sequence 3, Application PC/TUS9605800
GENERAL INFORMATION:
APPLICANT: Oravax, Inc.
TITLE OF INVENTION: MULTIMERIC, RECOMBINANT UREASE VACCINE
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US96/05800
FILING DATE: 23-APR-1996
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/431,041
FILING DATE: 28-APR-1995
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/568,122
FILING DATE: 06-DEC-1995
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Clark, Paul T.
REGISTRATION NUMBER: 30,162
REFERENCE/DOCKET NUMBER: 06132/020001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 566 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US96-05800-3

Query Match 4.6%; Score 136.5; DB 2; Length 566;
Best Local Similarity 19.1%; Pred. No. 2.2e-05;
Matches 121; Conservative 80; Mismatches 220; Indels 211; Gaps 29;

QY 17 LLIRGRIVNDQSFYADVHDEGLIKQIG-----NLIVPGGIHTIDAHGLM 64
DB 68 LIITNALIVDTGIYKADIGKDGKTAGIKGKGNKMDQGVKNLSVGPATEALAGEGLI 127

QY 65 VLPGGVDVHTRLOMPVLGMPADDFCQGTKAALAGG-TTILHDHVPDGTGVSLLAAEQW 123
DB 128 VTAGGIDTHIFISP-----QQIPTAFASGVTTMIGGTGPGADGTNA----- 169

QY 124 RERADSAACDYSLVHDIITRWHSIEELEALVKEGVNSFLVFMAYKDRCCSDSQMYE 183
DB 170 -----TTIPGRRLNKLWMLRA-AEEYSMN----- 192

QY 184 IFSIIRDLGALAQVHAENGDIVEEQKRLLELGTG-----PE--GHVLSHPPEV 231
DB 193 -----LGFLAKGNASNDASLADQ-----IEAGAIGFKIHEDWGTTPSAINHALDVADKY 241

QY 232 EAEAVYRVTIAKQANCPLYTVKVMKSGAADAIAQAKRRGVVVFGEPTASLGTDGSHYW 291
DB 242 DVQVAIHTDTL-NEAGC-----VEDTMAAIAIGRTWHTF-----HTEGAGGGH-- 282

QY 292 SKNKAFAAFVTSPPVNPDP-----TTADHLTCLSSGDLQVVGSAHCTFTTA-----Q 340
DB 283 ADPIIKVAGEHNILPASTNPTPTPTVNTAEHMDMLVCHHLDKSIKEDVQFADSRIRPQ 342

QY 341 KAVGKDN-----FALIPGTNG--IEERMSWVEKCVAS-----GKMDENE----- 379
DB 343 TIAAEDTLHDMGIFSTSSDSQAMGRVGEVITRTWOTADKNKEFGRLKEEGDNDNFERI 402

QY 380 --FVAVTSTNAKIFNYPKRGVAVGSDADLVINWPKATKITSARKTNLNVYEIFEGV 437
DB 403 KRYLSKYTIINPATANGISEVSGVGVKADLVLWSP-----AFFGV 444

QY 438 ECRGAPAVVISQGRVALEDGKMFVTPGAGRFVPRKTFPFVYKRI-----KARNELAEI 491
DB 445 K-----PNMIKGGFIALSQ-----MGDANASIPT--PQVYIREMFAHKGAKYD-ANI 491

QY 492 HGVPRGLYDGPVHEVM-----VPAKPGSGAPARASCPGKISVPPVRNLHQSGFSLSGSQ 545
DB 492 TFSQAAYDKGIELGRLERQVLPVK-----NC-----RNITKMDQFNDDTT 533

ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/920,095
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/431,041
FILING DATE: 28-APR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Clark, Paul T.
REGISTRATION NUMBER: 30,162
REFERENCE/DOCKET NUMBER: 06132/013001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-8906
TELEFAX: (617) 542-5070
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 566 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-920-095-3

Query Match 4.6%; Score 136.5; DB 2; Length 566;
Best Local Similarity 19.1%; Pred. No. 2.2e-05;
Matches 121; Conservative 80; Mismatches 220; Indels 211; Gaps 29;

QY 17 LLIRGRIVNDQSFYADVHDEGLIKQIG-----NLIVPGGIHTIDAHGLM 64
DB 68 LIITNALIVDTGIYKADIGKDGKTAGIKGKGNKMDQGVKNLSVGPATEALAGEGLI 127

QY 65 VLPGGVDVHTRLOMPVLGMPADDFCQGTKAALAGG-TTILHDHVPDGTGVSLLAAEQW 123
DB 128 VTAGGIDTHIFISP-----QQIPTAFASGVTTMIGGTGPGADGTNA----- 169

QY 124 RERADSAACDYSLVHDIITRWHSIEELEALVKEGVNSFLVFMAYKDRCCSDSQMYE 183
DB 170 -----TTIPGRRLNKLWMLRA-AEEYSMN----- 192

QY 184 IFSIIRDLGALAQVHAENGDIVEEQKRLLELGTG-----PE--GHVLSHPPEV 231
DB 193 -----LGFLAKGNASNDASLADQ-----IEAGAIGFKIHEDWGTTPSAINHALDVADKY 241

QY 232 EAEAVYRVTIAKQANCPLYTVKVMKSGAADAIAQAKRRGVVVFGEPTASLGTDGSHYW 291
DB 242 DVQVAIHTDTL-NEAGC-----VEDTMAAIAIGRTWHTF-----HTEGAGGGH-- 282

QY 292 SKNKAFAAFVTSPPVNPDP-----TTADHLTCLSSGDLQVVGSAHCTFTTA-----Q 340
DB 283 ADPIIKVAGEHNILPASTNPTPTPTVNTAEHMDMLVCHHLDKSIKEDVQFADSRIRPQ 342

QY 341 KAVGKDN-----FALIPGTNG--IEERMSWVEKCVAS-----GKMDENE----- 379
DB 343 TIAAEDTLHDMGIFSTSSDSQAMGRVGEVITRTWOTADKNKEFGRLKEEGDNDNFERI 402

QY 380 --FVAVTSTNAKIFNYPKRGVAVGSDADLVINWPKATKITSARKTNLNVYEIFEGV 437
DB 403 KRYLSKYTIINPATANGISEVSGVGVKADLVLWSP-----AFFGV 444

QY 438 ECRGAPAVVISQGRVALEDGKMFVTPGAGRFVPRKTFPFVYKRI-----KARNELAEI 491
DB 445 K-----PNMIKGGFIALSQ-----MGDANASIPT--PQVYIREMFAHKGAKYD-ANI 491

QY 492 HGVPRGLYDGPVHEVM-----VPAKPGSGAPARASCPGKISVPPVRNLHQSGFSLSGSQ 545
DB 492 TFSQAAYDKGIELGRLERQVLPVK-----NC-----RNITKMDQFNDDTT 533

[illegible]

RESULT 8
 US-08-467-822-21
 ; Sequence 21, Application US/08467822
 ; Patent No. 5843460
 ; GENERAL INFORMATION:
 ; APPLICANT: Labigne, Agnes
 ; APPLICANT: Sauerbaum, Sebastien
 ; APPLICANT: Ferrero, Richard L.
 ; APPLICANT: Thiberge, Jean-Michel
 ; TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS AGAINST
 ; TITLE OF INVENTION: HELICOBACTER INFECTION. POLYPEPTIDES FOR USE IN THE
 ; TITLE OF INVENTION: COMPOSITIONS, AND NUCLEIC ACID SEQUENCES ENCODING SAID
 ; TITLE OF INVENTION: POLYPEPTIDES
 ; NUMBER OF SEQUENCES: 44
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
 ; ADDRESSEE: Dunner
 ; STREET: 1300 I Street, N.W.
 ; CITY: Washington
 ; STATE: D.C.
 ; COUNTRY: USA
 ; ZIP: 20005-3315
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC Compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/467,822
 ; FILING DATE: 06-JUN-1995
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/447,177
 ; FILING DATE: 19-MAY-1995
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/432,697
 ; FILING DATE: 02-MAY-1995
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Meyers, Kenneth J.
 ; REGISTRATION NUMBER: 25,146
 ; REFERENCE/DOCKET NUMBER: 03495.0137-02000
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (202) 408-4000

TELEFAX: (202) 408-4400
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 569 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
FEATURE:
NAME/KEY: Protein
LOCATION: 1..569
OTHER INFORMATION: /note="URE B - FIGURE 3."
US-08-467-832-21

Query Match	4.6%;	Score 136.5;	DB 2;	Length 569;
Best Local Similarity	19.7%;	Pred. No. 2.3e-05;		
Matches 109;	Conservative 65;	Mismatches 196;	Indels 181;	Gaps

QY	1	MSFOGKKSI-----PRITSDLLITRGGRVINDOSFYADVHVEDGLIKQIGE----	47
Db	43	IKEGGKTTKDCMSQTNSSPSSVELDLVTNALIVDYTIYKADIGIKDGIAGIKAGNK 102	
QY	48	-----NLIVPGGIHTIDAHGLMVLPGGVDDVHTRLQMPVLGTMPTADDFCQGTKAALG 99	
Db	103	DMQDGVNNLNCVPATEALAAEGLVITAGGIDTHIFISP-----QQIPTAFAS 151	
QY	100	G-TTMIIDHVPDTCVSLAAAYEQWREADSAAACDYSLHVDITRWHESIKEEALVKE 158	
Db	152	GVTTMIGGTTGPDAGTNA-----TTTTPGRANLKSMLRA-AEE 188	
QY	159	KGVNSFLVFMAYKRCQCSQSMQYEI-----FSIIRDLGAL-AQVHAENGDIVEEQKR 211	
Db	189	YAMN-YLGLFA-KGVNSYEPSLRQDIEAGAIKRIKIHEDMGSTPAAIH-----232	
QY	212	LELGITPEGHVLSHPEVEAEAVYRAVYIAQANCPLYTVKVMKSGAADAIAQAKRG 271	
Db	233	-----HCLNVADEYDVQVAIHTDFL-NEAGC-----VEDTLEAIAGRT 269	
QY	272	VVVFGEPTIASLGDGSHYKSKWKAAPVTSPPVNPDP-----TTADHLTCLLSSG 324	
Db	270	IHTF-----HTEGAGGGH--APDVIKMAGEFNLPASTNPTPTFTKNTAEAHMDMLWCH 322	
QY	325	DLQVTSAHCTFTTA-----QKAVGKDN-----FALIPEGTNG-----IEERMSWVEKCA 371	
Db	323	HLDSIKEDYQFADSRIRPQTIAAEDQLHDMGIFSTSSDSQAMGRVGEVIRTTQWTADK 382	
QY	372	S-----GKWDENE-----FVATYSTNAKINFYPRKRVAGSDADLVINWPKAT 417	
Db	383	NKKEFGRLKEEKGDNDNFRIKRYISKYITNPQIANGISDIYGSVEVGKYADLVLWSP-----439	
QY	418	KIISAKTHNLNVEYNTIFEGVECRGAPAVVISQGRALED-----GKMFV 461	
Db	440	-----AFFGIK-----PNMIKGGFTALSQMGDANASIPTTPQPVYYREMF 480	
QY	462	TPGAGRFVPRKTF 474	
Db	481	HGKNKFDNTNITF 493	

```

RESULT          9
US-08-432-697-21
; Sequence 21, Application US/08432697
; Patent No. 6248330
;
; GENERAL INFORMATION:
; APPLICANT: Labigne, Agnes
; APPLICANT: Sauerbaum, Sebastian
; APPLICANT: Ferrero, Richard L.
; APPLICANT: Thiberge, Jean-Michel
; TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS
; TITLE OF INVENTION: HELICOBACTER COMPOSITIONS
; TITLE OF INVENTION: POLYPEPTIDES
;

```

Db	383	NKKEPRLKEEKDNDNFRIKRYISKYITNGIAHGISDYVGSVEVGKYADLVLWSP----	439
Qy	418	KIISAETHNLNVEYNIFGVCGRGAPAVVISQGRVALED-----	GKMFV 461
Db	440	-----AFFGIK-----PNMIKGGFIALSQMGDANASITPPQVYYREMG 480	
Qy	462	TPCAGRGVPVKTF 474	
Db	481	HGKNKFDTNITF 493	
 RESULT 10			
US-08-467-822-26			
; Sequence 26, Application US/08467822			
; Patent No. 5843460			
; GENERAL INFORMATION:			
; APPLICANT: Labigne, Agnes			
; APPLICANT: Sauerbaum, Sebastien			
; APPLICANT: Ferrero, Richard L.			
; APPLICANT: Thiberge, Jean-Michel			
; TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS AGAINST			
; HELICOBACTER INFECTION, POLYPEPTIDES FOR USE IN THE			
; TITLE OF INVENTION: COMPOSITIONS, AND NUCLEIC ACID SEQUENCES ENCODING SAID			
; POLYPEPTIDES			
; NUMBER OF SEQUENCES: 44			
; CORRESPONDENCE ADDRESS:			
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &			
; ADDRESSEE: Dunner			
; STREET: 1300 I Street, N.W.			
; CITY: Washington			
; STATE: D.C.			
; COUNTRY: USA			
; ZIP: 20005-3315			
; COMPUTER READABLE FORM:			
; MEDIUM TYPE: Floppy disk			
; COMPUTER: IBM PC compatible			
; OPERATING SYSTEM: PC-DOS/MS-DOS			
; SOFTWARE: Patentin Release #1.0, Version #1.30			
; CURRENT APPLICATION DATA:			
; APPLICATION NUMBER: US/08/467,822			
; FILING DATE: 06-JUN-1995			
; CLASSIFICATION: 435			
; PRIOR APPLICATION DATA:			
; APPLICATION NUMBER: US 08/447,177			
; FILING DATE: 19-MAY-1995			
; CLASSIFICATION: 435			
; PRIOR APPLICATION DATA:			
; APPLICATION NUMBER: US 08/432,697			
; FILING DATE: 02-MAY-1995			
; CLASSIFICATION: 435			
; ATTORNEY/AGENT INFORMATION:			
; NAME: Meyers, Kenneth J.			
; REGISTRATION NUMBER: 25,146			
; REFERENCE/DOCKET NUMBER: 03495.0137-02000			
; TELECOMMUNICATION INFORMATION:			
; TELEPHONE: (202) 408-4000			
; TELEFAX: (202) 408-4400			
; INFORMATION FOR SEQ ID NO: 26:			
; SEQUENCE CHARACTERISTICS:			
; LENGTH: 569 amino acids			
; TYPE: amino acid			
; STRANDEDNESS: single			
; TOPOLOGY: linear			
; MOLECULE TYPE: protein			
; US-08-467-822-26			
 Query Match 4.5%; Score 134.5; DB 2; Length 569;			
Best Local Similarity 19.0%; Pred. No. 3.7e-05;			
Matches 121; Conservative 78; Mismatches 218; Indels 219; Gaps 30;			
Qy	17	LLIRGRIVNDQSFYADVHVEDGLIKQIGE-----NLIVPGGIHTIDARGLM 64	
Db	17	LLIRGRIVNDQSFYADVHVEDGLIKQIGE-----NLIVPGGIHTIDARGLM 64	

```

Db 68 LIITNALIVDYTGKADIGIKDKGIAGICKGNKMDQGVKNLSVGPATEALAGEGLI 127
QY 65 VLPGGVDVHTRLOMPVLGTMTPADDFCQGTKAALAGG-TTMIIDHVPFDTGVSLLAAYEQW 123
Db 128 VTAGGIDTHIFISP-----QQIPTAFASGVTTMIGGGTGPADGTNA----- 169
QY 124 RERADSAACCDYSLHVDITRWHSIKEELEALVKEKGVNSFLVFWKDRQCSDSQMYE 183
Db 170 -----TTITPGRNKLWMLRA-AEYSNN----- 192
QY 184 IFSIIRDLGALAOVHAENGDIVEEQKRLLELGTG-----PE--GHVLSHPPEV 231
Db 193 -----LGFLAKGNASN-DASARDQ-----IEAGAIGFKIHEDWGTTPSAINHALDVADKY 241
QY 232 EAEAVYRAVTIAKOANCPLYYTKVMSKGAADAIQAQKRRGVVVFGEPIASLGTGDSHYW 291
Db 242 DVQVAIHDTL-NEAGC-----VEDTMAAIAGRTHMTF-----HTEGAGGHH-- 282
QY 292 SKNWAATAAFVTSPPVNDP-----TTADHLTCLSSGDLQVTSAGHCTFTA-----Q 340
Db 283 APDIKIVAGEHNILPASTNPTFTVTNTEAHEMDMLVCHHLKSIKEDVQFADSRIRPQ 342
QY 341 KAVGKDN-----FALIPEGTNG-----IEERMSVMWKCVCAS-----GKMDENE----- 379
Db 343 TIAAEDTLHDMGIFESITSSDSQAMGRVGEVITRTWTQADANKKEFGKKEEGDGNDFRI 402
QY 380 --FVAVTSTNAKIFNFPYPRKGRVAVGSDADLVINWPKATKIISAKTHNLNVEYNIFEGV 437
Db 403 KRYLSKTYINPAIAHGISVGVSEVGVKADVLVWSP-----AFFGV 444
QY 438 ECRGAPAVISQGRVALEDGKMEVTPCAGRFVPRKFTFPDVFYKRIKARNLAEIHG----- 493
Db 445 K-----PNMIKGGFIALSQ-----MGDANASIPT---POPVIYR-----EMFGHHGKAKY 487
QY 494 -----VPRGLYDGPVHEVM-----VPKPGSGAPARASCPCGKISVPPVRLNHSQGSFL 541
Db 488 DRNITFVQAAYDKIGKELGLERQVLPVK-----NC-----RNITKDKMQF 529
QY 542 SGQADHDHART-----AQKIMAPPGGRSNTLSL 571
Db 530 NDTTAHIEVNPETYHYFVDGKEVTSKPANKVSLAQL 565

```

RESULT 11

```

US-08-432-697-26
; Sequence 26, Application US/08432697
; Patent No. 6248330
; GENERAL INFORMATION:
; APPLICANT: Labigne, Agnes
; APPLICANT: Sauerbaum, Sebastien
; APPLICANT: Ferrero, Richard L.
; APPLICANT: Thibierge, Jean-Michel
; TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS AGAINST
; TITLE OF INVENTION: HELICOBACTER INFECTION, POLYPEPTIDES FOR USE IN THE
; COMPOSITION: COMPOSITIONS, AND NUCLEIC ACID SEQUENCES ENCODING SAID
; TITLE OF INVENTION: POLYPEPTIDES
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; ADDRESSEE: Dunner
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/432,697
; FILING DATE: 02-MAY-1995

```

```

; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Meyers, Kenneth J.
; REGISTRATION NUMBER: 25,146
; REFERENCE/DOCKET NUMBER: 03495.0137-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 408-4000
; TELEFAX: (202) 408-4400
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 569 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-432-697-26

```

```

Query Match 4.5%; Score 134.5; DB 4; Length 569;
Best Local Similarity 19.0%; Pred. No. 3.7e-05;
Matches 121; Conservative 78; Mismatches 218; Indels 219; Gaps 30;

QY 17 LLIRGRIVNDQSFYADVHVEDGLIKQIGE-----NLIVPGGIHTIDAHGLM 64
Db 68 LIITNALIVDYTGKADIGIKDKGIAGICKGNKMDQGVKNLSVGPATEALAGEGLI 127
QY 65 VLPGGVDVHTRLOMPVLGTMTPADDFCQGTKAALAGG-TTMIIDHVPFDTGVSLLAAYEQW 123
Db 128 VTAGGIDTHIFISP-----QQIPTAFASGVTTMIGGGTGPADGTNA----- 169
QY 124 RERADSAACCDYSLHVDITRWHSIKEELEALVKEKGVNSFLVFWKDRQCSDSQMYE 183
Db 170 -----TTITPGRNKLWMLRA-AEYSNN----- 192
QY 184 IFSIIRDLGALAOVHAENGDIVEEQKRLLELGTG-----PE--GHVLSHPPEV 231
Db 193 -----LGFLAKGNASN-DASARDQ-----IEAGAIGFKIHEDWGTTPSAINHALDVADKY 241
QY 232 EAEAVYRAVTIAKOANCPLYYTKVMSKGAADAIQAQKRRGVVVFGEPIASLGTGDSHYW 291
Db 242 DVQVAIHDTL-NEAGC-----VEDTMAAIAGRTHMTF-----HTEGAGGHH-- 282
QY 292 SKNWAATAAFVTSPPVNDP-----TTADHLTCLSSGDLQVTSAGHCTFTA-----Q 340
Db 283 APDIKIVAGEHNILPASTNPTFTVTNTEAHEMDMLVCHHLKSIKEDVQFADSRIRPQ 342
QY 341 KAVGKDN-----FALIPEGTNG-----IEERMSVMWKCVCAS-----GKMDENE----- 379
Db 343 TIAAEDTLHDMGIFESITSSDSQAMGRVGEVITRTWTQADANKKEFGKKEEGDGNDFRI 402
QY 380 --FVAVTSTNAKIFNFPYPRKGRVAVGSDADLVINWPKATKIISAKTHNLNVEYNIFEGV 437
Db 403 KRYLSKTYINPAIAHGISVGVSEVGVKADVLVWSP-----AFFGV 444
QY 438 ECRGAPAVISQGRVALEDGKMEVTPCAGRFVPRKFTFPDVFYKRIKARNLAEIHG----- 493
Db 445 K-----PNMIKGGFIALSQ-----MGDANASIPT---POPVIYR-----EMFGHHGKAKY 487
QY 494 -----VPRGLYDGPVHEVM-----VPKPGSGAPARASCPCGKISVPPVRLNHSQGSFL 541
Db 488 DRNITFVQAAYDKIGKELGLERQVLPVK-----NC-----RNITKDKMQF 529
QY 542 SGQADHDHART-----AQKIMAPPGGRSNTLSL 571
Db 530 NDTTAHIEVNPETYHYFVDGKEVTSKPANKVSLAQL 565

```

RESULT 12

```

US-08-467-822-25
; Sequence 25, Application US/08467822
; Patent No. 5843460
; GENERAL INFORMATION:
; APPLICANT: Labigne, Agnes

```

APPLICANT: Sauerbaum, Sebastien
APPLICANT: Ferrero, Richard L.
APPLICANT: Thiberge, Jean-Michel
TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS AGAINST
TITLE OF INVENTION: HELICOBACTER INFECTION, POLYPEPTIDES FOR USE IN THE
TITLE OF INVENTION: COMPOSITIONS, AND NUCLEIC ACID SEQUENCES ENCODING SAID
TITLE OF INVENTION: POLYPEPTIDES
NUMBER OF SEQUENCES: 44
CORRESPONDENCE ADDRESS:
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
ADDRESSEE: Dunner
STREET: 1300 I Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/467,822
FILING DATE: 06-JUN-1995
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/447,177
FILING DATE: 19-MAY-1995
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/432,697
FILING DATE: 02-MAY-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Meyers, Kenneth J.
REGISTRATION NUMBER: 25,146
REFERENCE/DOCKET NUMBER: 03495.0137-02000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 408-4000
TELEFAX: (202) 408-4400
INFORMATION FOR SEQ ID NO: 25:
SEQUENCE CHARACTERISTICS:
LENGTH: 840 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-467-822-25

Query Match 4.3%; Score 129; DB 2; Length 840;
Best Local Similarity 19.3%; Pred. No. 0.00028;
Matches 101; Conservative 58; Mismatches 181; Indels 182; Gaps 20;

QY 15 DRLIRGGTRVNDQS-----FYADVHVEDGLIKOIGE- 47
DB 311 DECVFGGKVRIDMGQSCGHPAISLDVITNAVIVDTYGIKADIGIKDGLIASIGKA 370
QY 48 -----NLIVFGGIHTIDAGLMVLPGVDVHTRLQMPVLGWTTPADDFCQGTAA 96
DB 371 GNPDMGVFSNMIGANTEVIAEGELIVTAGGIDCHIHICPOLVY-----EAI 420
QY 97 LAGTTHMLDHPDPTGVSVLLAAYEQWRERADSAACCDLSLHVITRWHSIKEELEALV 156
DB 421 SSGITLVGGGTGPAAGT-----RATT----- 442
QY 157 KEKGVNSFLVFMAYKDRQCSDSQWEIFSIIIRDLAGAQAQVHAENGDIIVEEQRLLELG 216
DB 443 -----CTSPQTMRLMLQSTDLLPLNFGFTGKSSSKPDELHEIKAG 485
QY 217 ITGPEGHVLSPHEE--VEAEAVYRANTIAKQANCPVY--TKVMSKG--AADAIAQAKRRG 271
DB 486 AMGLKLH-----EDWGSTPAIDNCLTIAEHHDITQIINHDTLNEAGFVHSEIAAFKGR 540

QY 272 VVVFGEPTASLGTGSHYWSKNWAKAAAFVTPSPVNPDPPTA-----DHLTCLSSG 324
DB 541 IHVY-----HSEGAGGH--APDIKVGCGIKNVLPSTNTPRLTSTNIDHLDMLWVCH 593
QY 325 DL-----QVTSAGHCTP---TTAQKAVGKDNFALI-----PEGTNGIEERMSMWVE--- 367
DB 594 HLDREIPEDVAFHSRIRKKTIAAEDVLHDIGAISISSDSQAMGRVGEVISRTWOTADK 653
QY 368 -----KCVASKMDNEF-----VAVTSTNAKIFNFPKGRVAVGSDADLYWNP 414
DB 654 NKAQTGPLKCDSS---DNDNFRIKRYIAKTYINPAIAHGISQVYVSGVEGKLADLVLMKP 710
QY 415 KATLIISAHTNLNVEYNEPEGVECRGAPAVVISOGRAVED 456
DB 711 S-----FFGYK-----PEWIKGGWVAVAD 730

RESULT 13
US-08-432-697-25
Sequence 25 Application US/08432697
Patent No. 6248330
GENERAL INFORMATION:
APPLICANT: Labigne, Agnes
APPLICANT: Sauerbaum, Sebastien
APPLICANT: Ferrero, Richard L.
APPLICANT: Thiberge, Jean-Michel
TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS AGAINST
TITLE OF INVENTION: HELICOBACTER INFECTION, POLYPEPTIDES FOR USE IN THE
TITLE OF INVENTION: COMPOSITIONS, AND NUCLEIC ACID SEQUENCES ENCODING SAID
TITLE OF INVENTION: POLYPEPTIDES
NUMBER OF SEQUENCES: 44
CORRESPONDENCE ADDRESS:
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
ADDRESSEE: Dunner
STREET: 1300 I Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/432,697
FILING DATE: 02-MAY-1995
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Meyers, Kenneth J.
REGISTRATION NUMBER: 25,146
REFERENCE/DOCKET NUMBER: 03495.0137-00000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 408-4000
TELEFAX: (202) 408-4400
INFORMATION FOR SEQ ID NO: 25:
SEQUENCE CHARACTERISTICS:
LENGTH: 840 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-432-697-25

Query Match 4.3%; Score 129; DB 4; Length 840;
Best Local Similarity 19.3%; Pred. No. 0.00028;
Matches 101; Conservative 58; Mismatches 181; Indels 182; Gaps 20;

QY 15 DRLIRGGTRVNDQS-----FYADVHVEDGLIKOIGE- 47
DB 311 DECVFGGKVRIDMGQSCGHPAISLDVITNAVIVDTYGIKADIGIKDGLIASIGKA 370

QY 48 -----NLVPGGIHTIDAHGLMVLPGVDVHTRLOMPVLGMPADDFCQGTAA 96
Db 371 GNPDIMNGVSNMIGANTVIAAGELIVTAGDICHIIYICPOLVY-----EAI 420
QY 97 LAGGTTMIDHVPDPDRGVSLAAAYEQWRERADSAACCDYSLHVDITRWHESIKEELEALV 156
Db 421 SSGITTLVGGGTGPAAGT-----RATT----- 442
QY 157 KEGVNSFLVMAYKDRCCDSQMSQYEIFSIIRDLGALAQAQVHAENGDIIVEEQKRLLLELG 216
Db 443 -----CTPSPOTMLMLQSTDDLPINFGTGKSSSKPDELHEIKAG 485
QY 217 ITGPEGHVLSHPEE--VEAEAVYRAVIAQAQNCPLVY-TKVMKSG--AADAIAQAQKRG 271
Db 486 AMGLKLLH-----EDWGSTPAADNCLTIAEHDDIQIINIHTDLINEAGFVHSHIAAFKGR 540
QY 272 VVVEGEPITASLTGDSHYSKWNKAAAFVSPVNPDPPTA-----DHLTCLSSG 324
Db 541 IHTY-----HSEGAGGH--APDIKVCIGIKNVLPSTNPTPLTSNTIDEHLDMLMVCH 593
QY 325 DL-----QVTGSAHCTF--TTAAQKAVGKDNFALI-----PECTNGIEERMSMWVE--- 367
Db 594 HLDREIPEDVAFHSIRKKTIAEDVLDHIGALSISDSQAMGRVGEVISRTWOTADK 653
QY 368 -----KCVASGKMDENEF-----VAVTSTNAKIFNFPKGRVAVGSDADLVINWP 414
Db 654 NKAQTGPLKCDSS--DNDNFRKRYIAKTYINPAIAHGISQYVGSVEVGLADLVLPK 710
QY 415 KATKIISAKTHNLNVEYNIFEGVEGECAPAVVISQGRVALED 456
Db 711 S-----FFGTK-----PEMVIKGMVAWAD 730

RESULT 14
5206163-3
; Patent No. 5206163
; APPLICANT: RENARD, ANDRE; DINA, DINO; MARTIAL, JOSEPH
; TITLE OF INVENTION: DNA ENCODING BOVINE DIARRHEA
; VIRUS PROTEIN
; NUMBER OF SEQUENCES: 3
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/550,816
; FILING DATE: 06-JUL-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 331,037
; FILING DATE: 29-MAR-1989
; APPLICATION NUMBER: 752,981
; FILING DATE: 08-JUL-1985
; SEQ ID NO:3
; LENGTH: 2616
5206163-3

Query Match 3.8%; Score 112.5; DB 6; Length 2616;
Best Local Similarity 20.7%; Pred. No. 0.12; Indels 99; Gaps 13;
Matches 72; Conservative 45; Mismatches 131; Indels 99; Gaps 13;
QY 125 ERASACCDYSLHVDITRWHESIK-----EELEALVKEGVNSFLVMAYKDRCCQ 176
Db 2075 ENQOSPGLHDLKLETFHTIAQPSLKHVTGYVTWQLEAGINRKAAGFL---EKNNGEV 2131
QY 177 SDSQWYEFSTIIRDLGALAQAHA-----ENGDIVEEQKRLLLELGITG 219
Db 2132 LDSEKHLVDIIRDLKGRKIRYETAIPTKNEKRDVSDMQAGDIVDEKKPR----- 2183
QY 220 PEGHVLSPHEVEAEV--YRAVIAQAQNCPLVTKVMKSGAADAIAQAQKRGVVVEGE 277
Db 2184 ----VIQPEAKTRLATKYNMVKQ---PVVIPYEGKTPLEKIFNKVRKEWDLFNE 2236
QY 278 PITASLTGDC-----SHYKSNKAAAFVTSPPVNPDPPTADHLTCL 320
Db 2237 PVAVSFDTKAWDTQVTRDLRLIGEIQYRYKEWHKFIDITIDHVMVEVPVITAD----- 2291

QY 321 LSSGDLQVTSAGHCTFTTAQKAVGKDNFALIPECTNGIEERMSVMWKECVASGKMDENEF 380
Db 2292 ---GEVYI-----RNGQSGSQPD-----TSAGNSMLNVLTMWYAFCESTGV----- 2330
QY 381 VAVTSTNAKIFNFPKGRVAV-GSDADLVINWPKATKIISAKTHN 426
Db 2331 -----PYKSFN---RVARIHCVGDDGFLITERGLGTCKICQORDAN 2367

RESULT 15
US-08-804-227C-5
; Sequence 5, Application US/08804227C
; Patent No. 5876991
; GENERAL INFORMATION:
; APPLICANT: DeHoff, Bradley S.
; APPLICANT: Kuhstoss, Stuart A.
; APPLICANT: Rostock, Paul R., Jr.
; APPLICANT: Sutton, Kimberly L.
; TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: THOMAS G. PLANT 1501
; STREET: LILLY CORPORATE CENTER
; CITY: INDIANAPOLIS
; STATE: IN
; COUNTRY: USA
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: ASCII(DOS) Text only
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/804,227C
; FILING DATE: February 21, 1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Plant, Thomas, G.
; REGISTRATION NUMBER: 35,784
; REFERENCE/DOCKET NUMBER: X-8231
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-276-2459
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1611 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-804-227C-5

Query Match 3.4%; Score 102.5; DB 2; Length 1611;
Best Local Similarity 21.5%; Pred. No. 0.56;
Matches 129; Conservative 73; Mismatches 236; Indels 161; Gaps 30;
QY 33 ADVHVEDGLKIOI-----GENLIVPGGIHTIDAHGLMVLPGVDVHTRLOMPVLGMPAD 88
Db 691 AVVALRAGLIRYLAGRAMAVALPAGEVEA-GLAKWP-GVEV-----AAVNG--PAST 741
QY 89 FCQGTKAALAGGTTM-----ILDHVFPDPTGVSLAAAYEQWRERADSAACCDYSLHVDITR 143
Db 742 VVSGDRRAVAVVAVQAEVGVOARLP-----VDYASH---SR 776
QY 144 WHESIKKELEALVKEGVNSFLVMAYKDRCCDSQMSQYEIFSIIRDLGALAQAQVHAENG 203
Db 777 HVEDLKELEKLVLSGIRSPRVPVCSVAGEQCEPFDAGYFNRNLRNRFESAVVVG 836
QY 204 IVEEQKRLLELGTGPEGH-VLSHPEVEAEAVYRAV-----TIQAQNCPLVTKVMK 258
Db 837 LLEBGRHREFIEV-----SAHPVLVHALQTEAADRSHVATGTLRRQDDSP-----H 883
QY 259 GAADIAQAQKRGVVVVEGEPI-----TASLGT-----DGSYWSKNKAAAFVTSPPVN-- 308

Db 884 RLLTSTAFAWAGATLTWDPALPPGHLTTLPTYPFNHHYWLDTPTTPTTOSPTDAW 943
QY 309 -----PDPTTADHLTCLL-----SSGDLQVTGSAHCTETTAQKA 342
Db 944 RYRVTKALTEESTPASSPSGHLLVTPPTPEGRITLGDRAAGALARQCATVERLVVDPVA 1003
QY 343 VGKDNEALIPECTNGIEERMSMVKEKCVASGKMDENEFVAVTSTNAAKIFNFYPRKGRVA 402
Db 1004 VGRD-----GLAARLGERWDGVLSSLGADERPLRHPALNRAVMGTTLAQAALD 1053
QY 403 VGSDADLVWNPATKIISAKTHNLNVEYNIEFEGVECRGAPAVVISQGR-VALEDGKMFV 461
Db 1054 AGCEARINAVTREAVAVSPSEVP-----RDAGQLWGLGRGIALEHPSLW- 1098
QY 462 TPCAGRFVPRKTFPDPFVYKRIKAR--NRLAEIHGVPRGLYDGPVHEVMVPAPKPGSGAPAR 519
Db 1099 ----GGLIDLPAVPD---ERAWARAVRRL-----VPHG-----EDQIAAR-ASGAYGR 1138
QY 520 ASCPGKISVPPV---RNLHQSGFSL--SGSQA-DDHIARRTAQKIMAPPGGRSNITSL 572
Db 1139 RLLP-----APPAASRRCTCTPSGTVLVTGGTGALGGHLARRLAR-----GGTGHVLVLT 1187

Search completed: July 30, 2001, 11:40:58
Job time: 282 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2001, 11:35:11 ; Search time 21.29 Seconds
(without alignments)
1628.789 Million cell updates/sec

Title: US-09-367-496-8
Perfect score: 2985
Sequence: 1 MSFQKKSIPRTSDRLIR.....RTAQKIMAPPGRSNTSLS 572

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 50623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_0601.*
1: /SID88/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SID88/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SID88/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SID88/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SID88/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SID88/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SID88/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SID88/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SID88/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SID88/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SID88/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SID88/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SID88/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SID88/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SID88/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SID88/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SID88/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SID88/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SID88/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SID88/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SID88/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SID88/gcgdata/geneseq/geneseq/AA2001.DAT.*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	2882	96.5	553	19 AAW68489	Human partial ULIP
2	2803	93.9	572	19 AAW68488	Mouse ULIP-4 prote
3	2325	77.9	572	19 AAW68486	Mouse ULIP-1 prote
4	2151	72.1	572	19 AAW68487	Mouse ULIP-3 prote
5	1044	35.0	683	21 AAG46568	Arabidopsis thalia
6	1011.5	33.9	485	17 AAW03025	Pseudomonas hydan
7	945	31.7	588	21 AAG46569	Arabidopsis thalia
8	901	30.2	576	21 AAG46570	Arabidopsis thalia
9	837	28.0	471	13 AAR29715	Heat stable hydan
10	826	27.7	460	16 AAR69874	B.thermoglucosidasi
11	691.5	23.2	457	16 AAR82837	Agrobacterium radi

12	684.5	22.9	457	17 AAW03023	Bacillus sp. hydan
13	684.5	22.9	457	17 AAW03024	Agrobacterium hyda
14	600	20.1	167	21 AAB58948	Breast and ovarian
15	534	17.9	281	21 AAG23637	Arabidopsis thalia
16	464	15.5	283	21 AAG23636	Arabidopsis thalia
17	464	15.5	284	21 AAG23635	Arabidopsis thalia
18	408	13.7	153	21 AAB41399	Human ORF ORF1163
19	396.5	13.3	458	21 AAB26143	Arthrobacter hydan
20	396.5	13.3	458	21 AAB26146	Arthrobacter hydan
21	395.5	13.2	458	21 AAB26141	Arthrobacter hydan
22	394.5	13.2	458	21 AAB26140	Arthrobacter hydan
23	394.5	13.2	458	21 AAB26144	Arthrobacter hydan
24	394.5	13.2	458	21 AAB26145	Arthrobacter hydan
25	390.5	13.1	458	21 AAB26142	Corynebacterium g1
26	242	8.4	447	22 AAB80122	Arthrobacterium g1
27	240	8.0	506	21 AAG26653	Arabidopsis thalia
28	240	8.0	512	21 AAG26652	Arabidopsis thalia
29	216	7.2	397	22 AAB50341	B. lactofermentum
30	206	6.9	409	21 AAG26654	Arabidopsis thalia
31	200.5	6.7	129	21 AAB40494	Human ORF ORF258
32	193	6.5	568	14 AAR36387	Urease gamma subun
33	178.5	6.0	169	21 AAB43021	Human pancreatic c
34	145	4.9	40	21 AAB54409	Cat flea HMT allan
35	143	4.8	384	21 AAB29608	Klebsiella aerogen
36	142	4.8	567	19 AAW37774	Klebsiella aerogen
37	142	4.8	567	21 AAY81825	Helicobacter pylori
38	137.5	4.6	749	19 AAW80599	Part of protein w1
39	136.5	4.6	569	11 AAR04580	Helicobacter felis
40	136.5	4.6	569	16 AAR74337	H. felis structura
41	136.5	4.6	569	17 AAW06730	H. pylori urease B
42	136.5	4.6	569	17 AAW07194	H. pylori GHPO 124
43	136.5	4.6	569	19 AAW98511	Urease A and B sub
44	134.5	4.5	806	16 AAR67371	H. pylori urease u
45	133.5	4.5	569	16 AAR67378	

ALIGNMENTS

RESULT 1	
AAW68489	
ID	AAW68489 standard; Protein: 553 AA.
AC	AAW68489;
DT	08-DEC-1998 (first entry)
XX	Human partial ULIP-4 protein.
DE	Human; Unc-33-like phospho-protein; ULIP; rat; neoplasm; tumourigenesis;
KW	neurodegenerative disorder; diagnosis.
XX	Homo sapiens.
XX	Key
FH	Location/Qualifiers
FT	Misc-difference 56
FT	/note= "encoded by TGA"
XX	FR2759701-AL.
PN	
XX	21-AUG-1998.
PD	
XX	19-FEB-1997; 97FR-0001961.
PF	
XX	19-FEB-1997; 97FR-0001961.
PR	
XX	(INRM : INSERM INST NAT SANTE & RECH MEDICALE.
PA	
XX	Aguera M, Belin MF, Byk T, Honnorat J, Kolattukudy P;
PI	Quach TT, Sobel A;
PI	
XX	WPI: 1998-449610/39.
DR	N-PSDB; AAV60818.
DR	

```

XX PT Mouse and human ULIP poly:peptide(s) - useful in detection of
XX PT para-neoplastic neurological syndromes
XX PS
XX PS Claim 1; Fig 12; 90pp; French.
XX CC This sequence represents a partial human Unc-33-like phospho-protein
XX CC (ULIP)-4. The coding sequence was isolated based on similarity to the
XX CC rat ULIP sequence. Proteins of the ULIP family or their corresponding
XX CC nucleic acids can be used in compositions for treating neurodegenerative
XX CC disorders and neoplasms, especially for para-neoplastic neurological
XX CC syndromes and/or for the early diagnosis of tumourigenesis.
XX SQ Sequence 553 AA;

Query Match 96.5%; Score 2882; DB 19; Length 553;
Best Local Similarity 99.8%; Pred. No. 1e-268;
Matches 552; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MSFQGGKSIPTSDRLIRGGRIVNDQSFYADVHVEDGLIKQIGENLIVPGGIHTIDA 60
Db 1 msfqqkksipritsdrlirggrivnddgsfyadvhvedgllkqigenlivpggixtida 60

QY 61 HGLMVLPGGVVHTRLQMPVLGMPVLTADDPCQGTAAALAGTMTILDHVPDPTGVSLAAY 120
Db 61 hglmvlpggvvhttrlqmpvlgmtaddpcqgtkaalagttmildhvpdptgvsllaay 120

QY 121 EQWRERADSAACCDYSLHVDITRWHESTKEELEALVKEGVNSFLVFMAYKDRQCSDSQ 180
Db 121 eqwreradaaaccdyslhvdi trwhesikeelealvkegvnsflvfmaykdrqc sdsq 180

QY 181 MYEFSIIRDLGALAQVHAENGDIVEEQKRLLELGITGPEGHVLSPHEVEAEAVYRAV 240
Db 181 myeifsiirdlgalaqvaengdi veeqkrllelgitgpeghvlshpeveaeav yrav 240

QY 241 TIAKQANCLYVTKVMSKGAADAIQAQRGVRVVFGEPTASLGTDSGHYWSKNWAKAAA 300
Db 241 tiakqancplyvtkvmskgaadaia qargvrvvfgeptaslgtdsgshywsknw akaaa 300

QY 301 FVTSPVPNPDPPTADHLTCLSSGDLQVTGSAHCTFTTAQKAVGKDNFALIPGNGIEE 360
Db 301 fvtspvpnpdpptadhl tclssgdlqvtgsahctfttaqkavgkdnfalipg ngiee 360

QY 361 RMSMWWEKCVASGKMDENEFVAVTSTNAKIFNFYPRKGRVAVGSDADLVINWPKATKII 420
Db 361 rmsmwwekcvasgkmden efvavtstnaakifnfyprkgrvavgsdadlv inwpkatkii 420

QY 421 SAKTHNLNVEYNIFEVCEGRGAPAVVISQGRVALEDGKMFVTPGAGREVPKRTPDFVYK 480
Db 421 sakt hnlnevynifevcegr gapavvisqgrvaledgkmfvtpgagrevpkr tpdfvyk 480

QY 481 RIKARNRLAEITHGVPRGLYDGPVHEVMVPAKPGSGAPARASCPGKISVPPVRLNHQSQFS 540
Db 481 rikarnrlaeithgvpr gllydgpvhvmvpa kpgsgaparascpgkisvppv rlnhqsqfs 540

QY 541 LSGSQADDDHIARR 553
Db 541 lsgsqaddhhiarr 553

```

RESULT 2

AAW68488

ID AAW68488 standard; Protein; 572 AA.

XX AC

XX AC AAW68488;

XX DT

XX DT 08-DEC-1998 (first entry)

XX DE

XX DE Mouse ULIP-4 protein.

XX KW

KW Mouse; Unc-33-like phospho-protein; ULIP; rat; neoplasm; tumourigenesis; neurodegenerative disorder; diagnosis.

XX OS Mus musculus.
 XX PN FR2759701-AL.
 XX PD 21-AUG-1998.
 XX PF 19-FEB-1997; 97FR-0001961.
 XX PR 19-FEB-1997; 97FR-0001961.
 XX PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.
 XX PI Aguera M, Belin MF, Byk T, Honnorat J, Kolattukudy P;
 XX PI Quach TT, Sobel A;
 XX DR WPI; 1998-449610/39.
 XX DR N-PSDB; AAV60817.
 XX PT Mouse and human ULIP poly:peptide(s) - useful in detection of
 XX PT para-neoplastic neurological syndromes
 XX PS Claim 1; Fig 11; 90pp; French.
 XX CC This sequence represents the mouse Unc-33-like phospho-protein (ULIP)-4.
 XX CC The coding sequence was isolated based on similarity to the rat ULIP
 XX CC sequence. Proteins of the ULIP family or their corresponding nucleic
 XX CC acids can be used in compositions for treating neurodegenerative
 XX CC disorders and neoplasms, especially for para-neoplastic neurological
 XX CC syndromes and/or for the early diagnosis of tumourigenesis.
 XX SQ Sequence 572 AA;

Query Match 93.9%; Score 2803; DB 19; Length 572;
 Best Local Similarity 92.7%; Pred. No. 4.5e-261;
 Matches 530; Conservative 22; Mismatches 20; Indels 0; Gaps 0;
 QY 1 MSFQGGKSIPTSDRLIRGGRIVNDQSFYADVHVEDGLIKQIGENLIVPGGIHTIDA 60
 Db 1 msfqqkksipritsdrlirggrivnddgsfyadvhvedgllkqigenlivpggiktida 60
 QY 61 HGLMVLPGGVVHTRLQMPVLGMPVLTADDPCQGTAAALAGTMTILDHVPDPTGVSLAAY 120
 Db 61 hglmvlpggvvhttrlqmpvlgmtaddpcqgtkaalagttmildhvpdpgvsllaay 120
 QY 121 EQWRERADSAACCDYSLHVDITRWHESTKEELEALVKEGVNSFLVFMAYKDRQCSDSQ 180
 Db 121 eqwreradaaaccdyslhvdi trwhesikeelealvkegvnsflvfmaykdrqc sdsq 180
 QY 181 MYEFSIIRDLGALAQVHAENGDIVEEQKRLLELGITGPEGHVLSPHEVEAEAVYRAV 240
 Db 181 myeifsiirdlgalaqvaengdi veeqkrllelgitgpeghvlshpeveaeav yrav 240
 QY 241 TIAKQANCLYVTKVMSKGAADAIQAQRGVRVVFGEPTASLGTDSGHYWSKNWAKAAA 300
 Db 241 tiakqancplyvtkvmskgaadma qargvrvvfgeptaslgtdsgshywsknw akaaa 300
 QY 301 FVTSPVPNPDPPTADHLTCLSSGDLQVTGSAHCTFTTAQKAVGKDNFALIPGNGIEE 360
 Db 301 fvtspvpnpdpptadhl tclssgdlqvtgsahctfttaqkavgkdnfalipg ngiee 360
 QY 361 RMSMWWEKCVASGKMDENEFVAVTSTNAKIFNFYPRKGRVAVGSDADLVINWPKATKII 420
 Db 361 rmsmwwekcvasgkmden efvavtstnaakifnfyprkgrvavgsdadlv inwpratkii 420
 QY 421 SAKTHNLNVEYNIFEVCEGRGAPAVVISQGRVALEDGKMFVTPGAGREVPKRTPDFVYK 480
 Db 421 sakt hnlnevynifevcegr gapavvisqgrvaledgkmfvtpgagrevpkr tpdfvyk 480
 QY 481 RIKARNRLAEITHGVPRGLYDGPVHEVMVPAKPGSGAPARASCPGKISVPPVRLNHQSQFS 540
 Db 481 rikarnrlaeithgvpr gllydgpvhvmvpa kpgsgaparascpgkisvppv rlnhqsqfs 540

QY 541 LSGSQADHIAARRTAQKIMAPPGRSNTSL 572
 Db 541 LSGSQADHIAARRTAQKIMAPPGRSNTSL 572

RESULT 3

AAW68486
 ID AAW68486 standard; Protein; 572 AA.

AC AAW68486;

DT 08-DEC-1998 (first entry)

XX Mouse ULIP-1 protein.

XX Mouse; Unc-33-like phospho-protein; ULIP; rat; neoplasm; tumorigenesis;
 KW neurodegenerative disorder; diagnosis.

XX Mus musculus.

XX FR2759701-AL.

XX 21-AUG-1998.

PF 19-FEB-1997; 97FR-0001961.

PR 19-FEB-1997; 97FR-0001961.

XX (INRM) INSERM INST NAT SANTE & RECH MEDICALE.

PI Aguera M, Belin MF, Byk T, Honnorat J, Kolattukudy P;

PI Quach TT, Sobel A;

XX WPI; 1998-449610/39.

XX N-PSDB; AAV60815.

PT Mouse and human ULIP poly:peptide(s) - useful in detection of

PT para-neoplastic neurological syndromes

PS Claim 1; Fig 9; 90pp; French.

XX This sequence represents the mouse Unc-33-like phospho-protein (ULIP)-1.
 CC The coding sequence was isolated based on similarity to the rat ULIP
 CC sequence. Proteins of the ULIP family or their corresponding nucleic
 CC acids can be used in compositions for treating neurodegenerative
 CC disorders and neoplasms, especially for para-neoplastic neurological
 CC syndromes and/or for the early diagnosis of tumorigenesis.

XX Sequence 572 AA;

Query Match 77.9%; Score 2325; DB 19; Length 572;

Best Local Similarity 75.3%; Pred. No. 5.2e-215;

Matches 430; Conservative 70; Mismatches 71; Indels 0; Gaps 0;

QY 1 MSFGKSTPRITSDLLIRGGRIVNDQSFVADHVEGLIKOIGENLIVPGIHTIDA 60

Db 1 msyggkknlpittdrlllkgkivnddgsfyadiymedgllkqiglenilvpgvktlea 60

QY 61 HGLMVLPGGVVDVHRLQMPVLGTMTPADDFCQGTKAALAGGTTMILDHVPFDTGVSLAAY 120

Db 61 hsrmlvpggidvhrfcmqdgmtsaddffqgtkaalaggttmildhvpepgtsllaaf 120

QY 121 EQWRERADSACCDYSLHVDITRWHESIKEEELALYKEGVNSFLVFMAYKRCQCSDSQ 180

Db 121 dqvrewadskscdcyslhvdtewhkgigeamealvkhgvsflvymafkdrfqtidsq 180

QY 181 MYEFSTIRDLGALAOVHAENGDI VEEOKRLELGTGTGEGHVLSPHEVEAEAYRAV 240

Db 181 iyevlsvirdigaiaqvhacngdiaeaqrdldlgtpeghvlrpeveaeavnrsl 240

QY 241 TIAQANCPPLYVTVMKSGAADAIAQAKRGVVVFGEPTASLGTDGSHYWSKNWAKAAA 300

Db 241 tianqncplyvtkmpksaaeviaqarkgtvygepitaslgtgshywsknwaaaa 300
 QY 301 FVTSPVPNPDPPTADHLTCLLSGDLQVTSAGHCTFTTAOKAVGKDNFALIPGTCNGIEE 360
 Db 301 fvtspplspdpptpdflnslscgdlqvtgsahctfntaqkavgkdnftlpegtngtee 360
 QY 361 RMSMVWEKCVASGKMDENEFVAVTSTNAKIFNFYPRKGRVAVGSDADLVINPKATKII 420
 Db 361 rmsvldwkavvtgkmdenqfvavtstnaakvfnlyprkgrisvgsdadlvwdpsvktl 420
 QY 421 SAKTHNLNVEYNIFEGVECRGAPAVVISQGRVALEDGKMFVTPGAGRFVPRKTFPDPVYK 480
 Db 421 sakthnsaleynifegmecrgsplvvisgqkivledgtllhvtegsryiprkpfdfvyk 480
 QY 481 RIKARNRLAEHGVPRGLYDGPVHEVMVPAKPGSGAPARASCPCGKISVPPVRLHQSGFS 540
 Db 481 rikarsrlaelrgvprglydgpvcvsvtpktvtpassaktspakqgappvrvnlhqsgfs 540
 QY 541 LSGSQADHIAARRTAQKIMAPPGRSNTSL 571
 Db 541 lsgsqadidnprttqrivappggranitsl 571

RESULT 4

AAW68487

ID AAW68487 standard; Protein; 572 AA.

AC AAW68487;

DT 08-DEC-1998 (first entry)

XX Mouse ULIP-3 protein.

XX Mouse; Unc-33-like phospho-protein; ULIP; rat; neoplasm; tumorigenesis;
 KW neurodegenerative disorder; diagnosis.

XX Mus musculus.

XX FR2759701-AL.

XX 21-AUG-1998.

XX 19-FEB-1997; 97FR-0001961.

XX 19-FEB-1997; 97FR-0001961.

XX (INRM) INSERM INST NAT SANTE & RECH MEDICALE.

PI Aguera M, Belin MF, Byk T, Honnorat J, Kolattukudy P;

PI Quach TT, Sobel A;

XX WPI; 1998-449610/39.

XX N-PSDB; AAV60816.

PT Mouse and human ULIP poly:peptide(s) - useful in detection of
 PT para-neoplastic neurological syndromes

PS Claim 1; Fig 10; 90pp; French.

XX This sequence represents the mouse Unc-33-like phospho-protein (ULIP)-3.
 CC The coding sequence was isolated based on similarity to the rat ULIP
 CC sequence. Proteins of the ULIP family or their corresponding nucleic
 CC acids can be used in compositions for treating neurodegenerative
 CC disorders and neoplasms, especially for para-neoplastic neurological
 CC syndromes and/or for the early diagnosis of tumorigenesis.

XX Sequence 572 AA;

Query Match

Best Loca: Similarity 72.1%; Score 2151; DB 19; Length 572;

Matches 394; Conservative 85; Mismatches 92; Indels 0; Gaps 0;

QY 1 MSFQKKSIPRITSDRLIRGRIVNDOSFYADVHVEDGLIKQIGENLIVPGIHTIDA 60
Db 1 mshqgkksiphtsdrllirgrindogsfyadvyleoglikqigenlivpgvktiea 60
QY 61 HGLMVLPGGVVHTRLQMPVLQMTFADDFCQCTKAALAGTMTIMLDHVPDPDGVSLAAAY 120
Db 61 ngmivpggidvntylqkpsqgmteadffgqkaalagttmiidhvvpepgsslltsf 120
QY 121 EOWRERADSAACDYSILHVDIIRWHSIEKEELEALVKEKGVNSFLVFMAYKDRQCQSDSQ 180
Db 121 ekwheadtkscddylhvdicwgydvreelevlvqdkgvnsfvmaykdklyqmsdsq 180
QY 181 MYEIFSIRDLGALCAQVHAENGDIIVEEQRLLELIGITGPEGHVLSHPEVEAEAVYRAV 240
Db 181 lyeatftklgavilvhaengdliaeqekrilemgitgpeghalsrpeeleaeavfrai 240
QY 241 TIAKANCPLYTVKMSKGAADIAQAKRGGVVVGEPISTASLGTDGSHYWSKNWAKAAA 300
Db 241 aiagrinpcvpyitkvmksaaadialarkkgplvfgelpiaaslgtdgthywsknwaaaa 300
QY 301 FVTSPPVNPDPPTDHLKCLSSGDLQVTGSAHCNFTTAQKAVGKDNALIPETNGTIEE 360
Db 301 fvtspplspdpptpylksliacgdlvtgsgchcypyslaqkavgkdnfllipegvngtee 360
QY 361 RMSWWEKCVASGKMDENEFVAVTSTNAKIFNFYPRKRVAVGSDADLVWNPVKATKII 420
Db 361 rmtvvwdkavatkndenqfvavtstnaakifnlyprkrgriavgsdadvlwdpdkmkti 420
QY 421 SAKTHNLNVEYNIFGVRCRGAAPVVISQGRVALEDGKMFVTPGAGRVPRKTFDFVYK 480
Db 421 takshksveynifegmechdsplvvisqgkivfedgnisvskmgtrfipkrkpfpehlyq 480
QY 481 RIKARNLAEIHGVPRLGLYDGVHVMVPAKPGSGAPARASCPGKISVPPVNLHQSFGS 540
Db 481 xvlrskvfglshvrgmgydpvyeatphaaapsaespskhqppplrlnhqsfnfs 540
QY 541 LSGSQADHIAIRTAQKIMAPPGGRSNTSL 571
Db 541 lsgsqadhdnprtrtgrhriavappgrsntsl 571
RESULT 5
AAG46568
ID AAG46568 standard; Protein; 683 AA.
XX AC AAG46568;
XX DT 18-OCT-2000 (first entry)
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 58600.
XX KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
OS Arabidopsis thaliana.
XX PN EP1033405-A2.
XX PD 06-SEP-2000.
XX PF 25-FEB-2000; 2000EP-0301439.
XX PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 13-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.

AAW03025 standard; Protein; 485 AA.
 AAW03025;
 31-JAN-1997 (first entry)
 Pseudomonas hydantoinase.
 Hydantoinase; hydantoin; d-N-carbamoyl-alpha-amino acid;
 intermediate; drug synthesis; penicillin; cephalosporin;
 antibiotic.
 Pseudomonas sp.
 WO9620275-A1.
 04-JUL-1996.
 26-DEC-1995; 95WO-JP02688.
 28-DEC-1994; 94JP-0336865.
 (KANF) KANEGAFUCHI KAGAKU KOGYO KK.
 Ikenaka Y, Nanba H, Takahashi S, Takano M, Yajima K;
 Yamada Y;
 WPI; 1996-321848/32.
 N-PSDB; AAT31258.
 Prodn. of D-N-carbamoyl amino acid from 5-substd. hydantoin - using
 a recombinant hydantoinase derived from a strain of Pseudomonas,
 Agrobacterium or Bacillus
 Claim 19; Page 36-40; 54pp; English.
 D-N-carbamoyl-alpha-amino acid is produced from a 5-substituted
 hydantoin by treatment with a hydantoinase expressed by a
 transformant microorganism carrying a vector containing DNA coding
 for the hydantoinase and derived from Bacillus sp. KNK245,
 Agrobacterium sp. KNK712 or Pseudomonas sp. KNK003A. The D-N-
 carbamoyl-alpha-amino acid can be used for the production of
 optically active alpha amino acids (especially D-phenylglycine and
 D-phydroxyphenylglycine) as intermediates for drug synthesis,
 especially for the production of semi-synthetic penicillin and
 cephalosporin antibiotics.
 Sequence 485 AA;
 Query Match 33.9%; Score 1011.5; DB 17; Length 485;
 Best Local Similarity 45.9%; Pred. No. 1.5e-88;
 Matches 210; Conservative 79; Mismatches 162; Indels 7; Gaps 5;
 18 LIRGRIVNDOSFYADVHVHVEDGLIKOIGENLIYVPGGIHIDAGLMVLPGGVVHRLQ 77
 4 vikgttiadvrsyeadillegkiaigrdl---qgdkivdaegaylipgldphthle 60
 78 MPVLGMPADDFCQGTAKAAGGTTMILDHVPDTGVSLAAYEQWRERADSAACCDYSL 137
 61 mpfmgtttaetwesgtfaalsgggttmvdfvpgp-agmlaafqwceraarqassdysl 119
 138 HVDITRWHESKELEALVKEGVNSFLVPMAYKDRQCSDSQMYEFTSIIRDIGALAAQV 197
 120 hmcvtgswskqlfedm-akvvergvntfkfmaykgaImvndemfasfgrcaalgalplv 178
 198 HAENGDIVEEQKRLLELGTGTPGHVLSHPEEVEAEAVYRAVTIAKQNCPLVTKVMS 257
 179 haengdivaslkqymegltgpeahysrpevegeatramiaadaagvpyvihvsc 238
 258 KGAADATQAQRGGVVVFGPEPTIASLTGDSHYSKWNKAAAFVTPSPVNPDPPTADHL 317
 239 eqahealrrarqgmrvygeplqlhllldeseykradwdeaarvmsapfr-dkshqds 297

QY 318 TCLSSGDIQVTSACHTFTTAQKAVGKDNFALIPETNGIERMSVWVKCVASGKMBE 377
 Db 298 wagliaalslqvvatdchcaftteqkrlgndftkipngtgledrisllwtygvktgrltp 357
 QY 378 NEFVAVTSTNAAKIFNYPKRGVAVGSDADLVVWNPATKIISAKTHNLNVEYNIFEV 437
 Db 358 nefavtstniagilniypqgailpgsdadlvvwdpakskkitasakslidynifegy 417
 QY 438 ECRGAPAVVISQGRVAL-EDGKMFVTPGAGRFVPRKTF 474
 Db 418 evtgmprytlrgevvwgeegsneprpgrgrfverpf 455
 RESULT 7
 AAG46569
 ID AAG46569 standard; Protein; 588 AA.
 XX AC AAG46569;
 XX DT 18-OCT-2000 (first entry)
 XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 58601.
 KW Protein identification; signal transduction pathway; metabolic pathway;
 KW hybridisation assay; genetic mapping; gene expression control; promoter;
 KW termination sequence.
 XX OS Arabidopsis thaliana.
 XX PN EP1033405-A2.
 XX PD 06-SEP-2000.
 XX PF 25-FEB-2000; 2000EP-0301439.
 XX PR 25-FEB-1999; 99US-0121825.
 PR 05-MAR-1999; 99US-0123180.
 PR 09-MAR-1999; 99US-0123548.
 PR 23-MAR-1999; 99US-0125788.
 PR 25-MAR-1999; 99US-0126284.
 PR 29-MAR-1999; 99US-0126785.
 PR 01-APR-1999; 99US-0127462.
 PR 06-APR-1999; 99US-0128234.
 PR 08-APR-1999; 99US-0128714.
 PR 16-APR-1999; 99US-0129845.
 PR 19-APR-1999; 99US-0130077.
 PR 21-APR-1999; 99US-0130449.
 PR 23-APR-1999; 99US-0130510.
 PR 28-APR-1999; 99US-0130891.
 PR 30-APR-1999; 99US-0131449.
 PR 30-APR-1999; 99US-0132048.
 PR 04-MAY-1999; 99US-0132407.
 PR 05-MAY-1999; 99US-0132484.
 PR 06-MAY-1999; 99US-0132485.
 PR 06-MAY-1999; 99US-0132486.
 PR 07-MAY-1999; 99US-0132487.
 PR 11-MAY-1999; 99US-0132863.
 PR 14-MAY-1999; 99US-0134218.
 PR 14-MAY-1999; 99US-0134219.
 PR 14-MAY-1999; 99US-0134221.
 PR 18-MAY-1999; 99US-0134370.
 PR 19-MAY-1999; 99US-0134941.
 PR 21-MAY-1999; 99US-0135124.
 PR 21-MAY-1999; 99US-0135353.
 PR 24-MAY-1999; 99US-0135629.
 PR 25-MAY-1999; 99US-0136021.
 PR 27-MAY-1999; 99US-0136392.
 PR 28-MAY-1999; 99US-0136782.
 PR 01-JUN-1999; 99US-0137222.
 PR 03-JUN-1999; 99US-0137528.

PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 21-JUN-1999; 99US-0139763.
PR 22-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144337.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.

PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 08-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 23-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 26-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 22-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 31.7%; Score 945; DB 21; Length 588;
Best Local Similarity 39.4%; Pred. No. 5,1e-82;


```

PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 28-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.

```

```

PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159384.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.
PR 29-OCT-1999; 99US-0162143.

Query Match 30.2%; Score 901; DB 21; Length 576;
Best Local Similarity 38.9%; Pred. No. 8.6e-78;
Matches 195; Conservative 86; Mismatches 178; Indels 42; Gaps 10;

QY 78 MEVLGMPADDFCQGTKAALAGGTTMILDHVPDPGVSLAAYEOWRERADSAACCDYSL 137
Db 1 mefmtetiddffsgaaaggttmldhfvipvng-nlvagfeay-eknsrescmdygf 58
QY 138 HVDITRWHESKEELEALVKEGVNSFLVFMAYKDRCCSQSOMYEIFSIIRDGALAOV 197
Db 59 hmaatkdegvsrdmelmivkeginsfkflayksglmvtdldllleglkrcksigalamv 118
QY 198 HAENGDIVEEQKRLELGITGPEGHVLSHPEEVEAEAVYRAVTTAKAANCPLYVTKWS 257
Db 119 haengdavfegqkrmiegitgpeghalsrppvlegatarairlarfintplyvvhms 178
QY 258 KGAADAIQAQKRGRGVVFEGETASIGTDSHYWSKNKAKAAAFVTSPPVNPDPPTADH- 316
Db 179 vdamdeiakarksgkvigepvvsllldhldhwpdffiaskyvmppirp-vgbhg 234
QY 317 --LTCLLSSGDLQVTSAGHCTFTTAQKAVGKDNFALIPETNGIIEERMSVMWEKCVASGK 374
Db 235 kalqdaistgilvgttdhctfnstgkalglldfrirpvgngleermhliwdtmvesgg 294
QY 375 MDENEFVAVTSTNAKIFNFYPRKGRVAVGSDADLVINPNKATKIISAKTHNLNVEYNIF 434
Db 295 lsatdyvritstecarifniyprkgailagsdadliilpnssyelskskshrsrdtnvy 354
QY 435 EGVECRGAPAVVISQGRVALEDGKMFVTPGAGRFVPRKFTFPDFVVKRIK----- 483
Db 355 egrrgkgvteviaggrivweneelkvprsgkytempf-sylfdeeketeckllslin 413
QY 484 ---ARNRLAEIRGVRGLYDGPVHEVMVPAKPGSCGAPARASC-----PGKISVPP---- 530
Db 414 icvsnnqafeleaqaqkffhggdnk--spankgrrrrreamtittislrksntrlppevnr 471
QY 531 ---VRNLHOSGFSLSGSQADD 548
Db 472 vlyvrnl---pfnitseemyd 489

RESULT 9
AAR29715
ID AAR29715 standard; Protein; 471 AA.
XX
AC AAR29715;
XX
DT 05-MAY-1993 (first entry)
XX
DE Heat stable hydantoinase protein.

```

XX Hydantoinase; E.coli; heat; stable; thermal.
 XX Bacillus stearothermophilus.
 OS JP04325093-A.
 PN 13-NOV-1992.
 PD 23-APR-1991; 91JP-0117802.
 PF 23-APR-1991; 91JP-0117802.
 PR (NIPS) NIPPON SODA CO.
 XX WPI; 1992-428828/52.
 DR N-PSDB; AAQ31987.
 XX Heat-stable hydantoinase (I) gene - has specified base and
 PT aminoacid sequence and is prepd. from E. coli
 XX Claim 2; Page 5-7; 8pp; Japanese.
 XX This sequence represents a heat-stable hydantoinase protein. The
 CC DNA encoding this peptide can be used to transform E.coli such
 CC that the protein is produced stably by the microbe.
 XX SQ Sequence 471 AA;

Query Match 28.08; Score 837; DB 13; Length 471;
 Best Local Similarity 40.08; Pred. No. 9e-72;
 Matches 183; Conservative 79; Mismatches 192; Indels 4; Gaps 4;
 QY 18 LIRGRIVNDQSFYADVHVEDGLIKQIGENLIVPGGIHTIDAHGLMVLPGGVVHTRLQ 77
 Db 4 iikngtvtatdiyeaddllidqgkiavgrnl-desgaevidatgcyvfpqgigphthld 62
 QY 78 MPVLGMPADDFCOGTAALAGGTTMILDHVFPDGTGVSLLAAYEQWRERADSAACCDYSL 137
 Db 63 mpfggtvkddsfesgtiaaafgggtttidfcitnkgelplkkaletwhnkagkavidygf 122
 QY 138 HVDITRWHESIKELEALVKEGVNSFLVFMAYKDRQCQSDQSMYEFISIRDLGALAQV 197
 Db 123 hlmseitddvleelpkvieegitsfkvmaykdvfgaddgtlyrtlvaakelgalvmv 182
 QY 198 HAENGDIVVEEQKRLLEIGITGPEGHVLSHPPEVEAEAVYRAVVTIAKOANCPLYVTKVMS 257
 Db 183 haengdivdyitkkaledgtdpiyhaltrppelegeatgracqltelagsqlvyhvc 242
 QY 258 KGADATAQAQRKRGVVVEGPEITASLGTDSGHVSKWNAKAAAFVTPPVNPDPTTADHL 317
 Db 243 aqavekiaearnkglvnmgetcpqylvdqsyilekpnf-egakyvwsppir-ekwhqevl 300
 QY 318 TCLSSGDLQVTSAGHCTFT-TAQKAVGKDNFALIPETNGIEERMSVMWEKCVASGKMD 376
 Db 301 wnalngqltldsdqcsfdkqgelgrdftkipnggpliedrvsilfsegvkkgrit 360
 QY 377 ENFVAVTSTNAAKIFNFPYPRKGVAVGSDADLVINWPKATKIISAKTHNLNVEYNIFEG 436
 Db 361 lngfdivstriaklflgfpkgtiavgadadlvifdptvervisaeethmavdynpfe 420
 QY 437 VECCGAPAVISQGRVALEDGKMFVETPGAGRFVPRKTF 474
 Db 421 mkvtgepvsvlcrgefvrldkqfvgkpgygyvkraky 458

RESULT 10

AAR69874

ID AAR69874 standard; Protein; 460 AA.

XX AC AAR69874;

XX

DT 27-OCT-1995 (first entry)
 XX B. thermoglucodasius D-hydantoinase with mutated C-terminal region.
 DE D-hydantoinase; dihydropyrimidinase; mutein; variant; mutation;
 KW D-N-carbamoyl-alpha-amino acid; penicillin; cephalosporin;
 KW production; racemisation.
 XX OS Bacillus thermoglucodasius.
 XX DE4328829-A.
 PN 02-MAR-1995.
 PD 27-AUG-1993; 93DE-4328829.
 PF 27-AUG-1993; 93DE-4328829.
 XX 27-AUG-1993; 93DE-4328829.
 PR (BOEF) BOEHRINGER MANNHEIM GMBH.
 PA Burtscher H, Lang G, Popp F;
 PI WPI; 1995-099331/14.
 XX N-PSDB; AAQ80565.
 DR New highly active and stable D-hydantoinase enzyme - isolated
 CC from Bacillus thermoglucodasius, useful in the prodn. of
 CC D-N-carbamoyl-alpha-amino acid for synthesis of, e.g. penicillin
 XX Claim 1; Page 5-6; 8pp; German.
 XX The wild-type D-hydantoinase gene was isolated from Bacillus
 CC thermoglucodasius using the amplification primers Hyd1 and Hyd2 (see
 CC AAQ80566 and AAQ80567). A single-base deletion in the HindIII site (from
 CC AACCTT to AACCT) was then introduced to give a coding region (AAQ80565)
 CC that encoded a D-hydantoinase (AAR69874) that was shorter than the
 CC wild-type protein with a different C-terminal sequence (see Features
 CC Table). The mutant enzyme was found to have improved heat stability
 CC and higher activity compared to known hydantoinases.
 XX SQ Sequence 460 AA;

Query Match 27.78; Score 826; DB 16; Length 460;
 Best Local Similarity 39.78; Pred. No. 1e-70;
 Matches 182; Conservative 78; Mismatches 194; Indels 4; Gaps 4;
 QY 18 LIRGRIVNDQSFYADVHVEDGLIKQIGENLIVPGGIHTIDAHGLMVLPGGVVHTRLQ 77
 Db 4 iikngtvtatdiyeaddllidqgkiavgrnl-desgaevidatgcyvfpqgigshthld 62
 QY 78 MPVLGMPADDFCOGTAALAGGTTMILDHVFPDGTGVSLLAAYEQWRERADSAACCDYSL 137
 Db 63 mpfggtvkddsfesgtiaaafgggtttidfcitnkgelplkkaletwhnkagkavidygf 122
 QY 138 HVDITRWHESIKELEALVKEGVNSFLVFMAYKDRQCQSDQSMYEFISIRDLGALAQV 197
 Db 123 hlmseitddvleelpkvieegitsfkvmaykdvfgaddgtlyrtlvaakelgalvmv 182
 QY 198 HAENGDIVVEEQKRLLEIGITGPEGHVLSHPPEVEAEAVYRAVVTIAKOANCPLYVTKVMS 257
 Db 183 haengdivdyitkkaledgtdpiyhaltrppelegeatgracqltelagsqlvyhvc 242
 QY 258 KGADATAQAQRKRGVVVEGPEITASLGTDSGHVSKWNAKAAAFVTPPVNPDPTTADHL 317
 Db 243 aqavekiaearnkglvnmgetcpqylvdqsyilekpnf-egakyvwsppir-ekwhqevl 300
 QY 318 TCLSSGDLQVTSAGHCTFT-TAQKAVGKDNFALIPETNGIEERMSVMWEKCVASGKMD 376
 Db 301 wnalngqltldsdqcsfdkqgelgrdftkipnggpliedrvsilfsegvkkgrit 360
 QY 377 ENFVAVTSTNAAKIFNFPYPRKGVAVGSDADLVINWPKATKIISAKTHNLNVEYNIFEG 436
 Db 361 lngfdivstriaklflgfpkgtiavgadadlvifdptvervisaeethmavdynpfe 420

[illegible]

Db 3 iikngtvtadgisradigikdgkktqig-----gaigpaertidaagryvfpqgidv 56
 QY 73 HTRLOMPVLGMPADDFCOGQTKAALAGGTTMILDHVPDTGVSLLAAYEQWRERADSAAC 132
 Db 57 hthvetsfntqsadtfatavaacggtttivdfcqqdgrghslaepvpkwdmaggksa 116
 QY 133 CDYSLHVDITRWHESKEELEALVKEKGVNSFLVFMAYKDRQCQSDSQMYEIFSIIIRDLG 192
 Db 117 idygyhiivldptdsvieeelevl-pdlgitsfkvmayrgmmidtdtllktldkavktg 175
 QY 193 ALAQVHAENGDIIVEEQKLELLELGTGPGCHVLSHPPEVEAEAVRANVTIAKOANCPLV 252
 Db 176 slvmvhaengdaadyldrkfvaegktapiyhalssrprveaeataralalaevinaplyi 235
 QY 253 TKVMSKGAADAIQAQRGRGVVFGEPITASLGTGDSHYSKNA-----KAAAFVTSPPV 307
 Db 236 vhtceesleevmkrakgrvraletcthyt-----yltkedlerpdegakvftppa 289
 QY 308 NPDPTADH--LTCLLSSGDLQVTSAGHCTFT-TAQKAVGKDNFALIPBGTNGIERMSM 364
 Db 290 r---akkdhvlnalrngvftvssdhcswlfkghkdrndfraipngapgveerlmm 346
 QY 365 VWEKCVASGKMDENEFVAVTSTNAKIFNFYPRKGRVAVGSDADLVINPKATKIISAKT 424
 Db 347 vygg-vnegrisltqfvelvatrpakvfgmpqkgtiavgsdadivlwdpeaemvieqta 405
 QY 425 HNLNVEYNIFEVGEGRGAPAVVISQGRVALEDGKMFVTPGAGRFPVPRKTF 474
 Db 406 mnamdyssyeghkvkvpktvllrgkvlvdegsvygeptdgkfkkrky 455

RESULT 13

AAW03024
 ID AAW03024 standard; Protein; 457 AA.

XX AC AAW03024;

XX DT 31-JAN-1997 (first entry)

XX DE Agrobacterium hydanotoinase.

XX KW Hydanotoinase; hydanotoin; d-N-carbamoyl-alpha-amino acid;
 KW intermediate; drug synthesis; penicillin; cephalosporin;
 XX antibiotic.

XX OS Agrobacterium sp.

XX FN WO9620275-A1.

XX PD 04-JUL-1996.

XX PF 26-DEC-1995; 95WO-JP02688.

XX PR 28-DEC-1994; 94JP-0326865.

XX PA (KANF) KANEGAFUCHI KAGAKU KOGYO KK.

XX PI Ikenaka Y, Nanba H, Takahashi S, Takano M, Yajima K;
 PI Yamada Y;

XX XX WPI; 1996-321848/32.

XX DR N-PSDB; AAT31257.

XX PT Prodn. of D-N-carbamoyl amino acid from 5-substd. hydanotoin - using
 PT a recombinant hydanotoinase derived from a strain of Pseudomonas,
 PT Agrobacterium or Bacillus

XX PS Claim 15; Page 30-34; 54pp; English.

XX CC D-N-carbamoyl-alpha-amino acid is produced from a 5-substituted
 CC hydanotoin by treatment with a hydanotoinase expressed by a
 CC transformant microorganism carrying a vector containing DNA coding

CC for the hydanotoinase and derived from Bacillus sp. KNK245,
 CC Agrobacterium sp. KNK712 or Pseudomonas sp. KNK003A. The D-N-
 CC carbamoyl-alpha-amino acid can be used for the production of
 CC optically active alpha amino acids (especially D-phenylglycine and D-
 CC p-hydroxyphenylglycine) as intermediates for drug synthesis,
 CC especially for the production of semi-synthetic penicillin and
 CC cephalosporin antibiotics.

XX SQ Sequence 457 AA;

Query Match 22.9%; Score 684.5; DB 17; Length 457;
 Best Local Similarity 34.3%; Pred. No. 4.2e-57;
 Matches 161; Conservative 83; Mismatches 197; Indels 29; Gaps 9;

QY 17 LIIRGRIVNDQSFYADVHVEDGLIKQIGENLIIVPGI-----HTIDAHLMLVPGVDV 72
 Db 3 iikngtvtadgisradigikdgkktqig-----gaigpaertidaagryvfpqgidv 56
 QY 73 HTRLOMPVLGMPADDFCOGQTKAALAGGTTMILDHVPDTGVSLLAAYEQWRERADSAAC 132
 Db 57 hthvetsfntqsadtfatavaacggtttivdfcqqdgrghslaepvpkwdmaggksa 116
 QY 133 CDYSLHVDITRWHESKEELEALVKEKGVNSFLVFMAYKDRQCQSDSQMYEIFSIIIRDLG 192
 Db 117 idygyhiivldptdsvieeelevl-pdlgitsfkvmayrgmmidtdtllktldkavktg 175
 QY 193 ALAQVHAENGDIIVEEQKLELLELGTGPGCHVLSHPPEVEAEAVRANVTIAKOANCPLV 252
 Db 176 slvmvhaengdaadyldrkfvaegktapiyhalssrprveaeataralalaevinaplyi 235
 QY 253 TKVMSKGAADAIQAQRGRGVVFGEPITASLGTGDSHYSKNA-----KAAAFVTSPPV 307
 Db 236 vhtceesleevmkrakgrvraletcthyt-----yltkedlerpdegakvftppa 289
 QY 308 NPDPTADH--LTCLLSSGDLQVTSAGHCTFT-TAQKAVGKDNFALIPBGTNGIERMSM 364
 Db 290 r---akkdhvlnalrngvftvssdhcswlfkghkdrndfraipngapgveerlmm 346
 QY 365 VWEKCVASGKMDENEFVAVTSTNAKIFNFYPRKGRVAVGSDADLVINPKATKIISAKT 424
 Db 347 vygg-vnegrisltqfvelvatrpakvfgmpqkgtiavgsdadivlwdpeaemvieqta 405
 QY 425 HNLNVEYNIFEVGEGRGAPAVVISQGRVALEDGKMFVTPGAGRFPVPRKTF 474
 Db 406 mnamdyssyeghkvkvpktvllrgkvlvdegsvygeptdgkfkkrky 455

RESULT 14

AA58948

ID AAB58948 standard; Protein; 167 AA.

XX AC AAB58948;

XX DT 27-MAR-2001 (first entry)

XX DE Breast and ovarian cancer associated antigen protein sequence SEQ ID 656.

XX KW Human; breast cancer; ovarian cancer; cytostatic; immunosuppressive;
 KW neoplastic; neuroprotective; antiviral; antiallergic; hepatotropic;
 KW antidiabetic; antiinflammatory; antitumor; vulnerary; anticonvulsant;
 KW antibacterial; antifungal; antiparasitic; cardiac; immune disorder;
 KW Addison's disease; allergy; autoimmune haemolytic anaemia;
 KW autoimmune thyroiditis; diabetes mellitus; Crohn's disease;
 KW multiple sclerosis; rheumatoid arthritis; ulcerative colitis;
 KW cardiovascular disorder; wound healing; neurological disease.

XX OS Homo sapiens.

XX PN WO200055173-A1.

XX XX 21-SEP-2000.

XX PD

Tue Jul 31 13:08:31 2001

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model
Run on: July 30, 2001, 11:40:36 ; Search time 21.48 seconds
(without alignments)
1614.382 Million cell updates/sec

Title: US-09-367-496-8
Perfect score: 572
Sequence: 1 MSFGKKSIPRTSDLLR.....RTAKTMAPGGRSNITSLS 572

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0
Searched: 412676 seqs, 60623988 residues
Word size : 0
Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Listing first 45 summaries

- Database : A_Geneseq_0601.*
- 1: /SID88/gcgdata/geneseq/geneseq/AA1980.DAT.*
 - 2: /SID88/gcgdata/geneseq/geneseq/AA1981.DAT.*
 - 3: /SID88/gcgdata/geneseq/geneseq/AA1982.DAT.*
 - 4: /SID88/gcgdata/geneseq/geneseq/AA1983.DAT.*
 - 5: /SID88/gcgdata/geneseq/geneseq/AA1984.DAT.*
 - 6: /SID88/gcgdata/geneseq/geneseq/AA1985.DAT.*
 - 7: /SID88/gcgdata/geneseq/geneseq/AA1986.DAT.*
 - 8: /SID88/gcgdata/geneseq/geneseq/AA1987.DAT.*
 - 9: /SID88/gcgdata/geneseq/geneseq/AA1988.DAT.*
 - 10: /SID88/gcgdata/geneseq/geneseq/AA1989.DAT.*
 - 11: /SID88/gcgdata/geneseq/geneseq/AA1990.DAT.*
 - 12: /SID88/gcgdata/geneseq/geneseq/AA1991.DAT.*
 - 13: /SID88/gcgdata/geneseq/geneseq/AA1992.DAT.*
 - 14: /SID88/gcgdata/geneseq/geneseq/AA1993.DAT.*
 - 15: /SID88/gcgdata/geneseq/geneseq/AA1994.DAT.*
 - 16: /SID88/gcgdata/geneseq/geneseq/AA1995.DAT.*
 - 17: /SID88/gcgdata/geneseq/geneseq/AA1996.DAT.*
 - 18: /SID88/gcgdata/geneseq/geneseq/AA1997.DAT.*
 - 19: /SID88/gcgdata/geneseq/geneseq/AA1998.DAT.*
 - 20: /SID88/gcgdata/geneseq/geneseq/AA1999.DAT.*
 - 21: /SID88/gcgdata/geneseq/geneseq/AA2000.DAT.*
 - 22: /SID88/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	497	86.9	553	19 AAW68489	Human partial ULIP
2	50	8.7	572	19 AAW68488	Mouse ULIP-4 prote
3	31	5.4	572	19 AAW68486	Mouse ULIP-1 prote
4	18	3.1	572	19 AAW68487	Mouse ULIP-3 prote
5	15	2.6	167	21 AAB58948	Breast and ovarian
6	13	2.3	40	21 AAB54409	Human pancreatic c
7	12	2.1	27	19 AAW69259	DMS protein fragme
8	11	1.9	18	19 AAW69260	DMS protein fragme
9	11	1.9	20	19 AAW69262	DMS protein fragme
10	10	1.7	460	16 AAR69874	B.thermoglucodasiu
11	10	1.7	485	17 AAW03025	Pseudomonas hydant

12	10	1.7	576	21 AAG46570	Arabidopsis thalia
13	10	1.7	588	21 AAG46569	Arabidopsis thalia
14	10	1.7	683	21 AAG46568	Arabidopsis thalia
15	9	1.6	283	21 AAG23636	Arabidopsis thalia
16	9	1.6	284	21 AAG23635	Arabidopsis thalia
17	8	1.4	169	21 AAB43021	Agrobacterium radi
18	8	1.4	457	16 AAR82837	Streptomyces chole
19	8	1.4	526	15 AAR60325	Cholesterol oxidas
20	8	1.4	546	17 AAW08932	Cholesterol oxidas
21	8	1.4	546	17 AAW08933	Cholesterol oxidas
22	8	1.4	546	17 AAW08934	Cholesterol oxidas
23	8	1.4	546	17 AAW08935	Cholesterol oxidas
24	8	1.4	546	17 AAW08936	Cholesterol oxidas
25	8	1.4	546	17 AAW08930	Wild type choleste
26	3	1.4	546	17 AAW08931	Human aggreacan deg
27	3	1.4	837	20 AAW75425	Human PRO1563 (UNO
28	3	1.4	837	21 AAY99429	Protein of the inv
29	8	1.4	837	22 AAB66178	L-galactono-1,4-la
30	7	1.2	15	18 AAW10617	Secreted protein o
31	7	1.2	55	19 AAW69340	Arabidopsis thalia
32	7	1.2	90	21 AAG33074	Arabidopsis thalia
33	7	1.2	90	21 AAG48928	Arabidopsis thalia
34	7	1.2	91	21 AAG33073	Arabidopsis thalia
35	7	1.2	91	21 AAG48927	Arabidopsis thalia
36	7	1.2	93	22 AAB80092	C. histolyticum CH
37	7	1.2	110	19 AAW63722	Arabidopsis thalia
38	7	1.2	121	21 AAG24039	Arabidopsis thalia
39	7	1.2	124	18 AAW20410	H. pylori cytoplas
40	7	1.2	124	18 AAW24654	H. pylori cytoplas
41	7	1.2	125	15 AAR62920	Human cytomagalovi
42	7	1.2	125	15 AAR62918	Human cytomagalovi
43	7	1.2	126	14 AAR36700	Flagellin (amino a
44	7	1.2	134	21 AAG24038	Arabidopsis thalia
45	7	1.2	139	21 AAY43865	Heavy chain (VH) 9

ALIGNMENTS

RESULT 1
AAW68489
ID AAW68489 standard; Protein: 553 AA.
XX
AC AAW68489;
XX
DT 08-DEC-1998 (first entry)
XX
DE Human partial ULIP-4 protein.
XX
KW Human; Unc-33-like phospho-protein; ULIP; rat; neoplasm; tumourigenesis;
neurodegenerative disorder; diagnosis.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Misc-difference 56 /note- "encoded by TGA"
FT
XX
PN FR2759701-A1.
XX
PD 21-AUG-1998.
XX
PF 19-FEB-1997; 97FR-0001961.
XX
PR 19-FEB-1997; 97FR-0001961.
XX
PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.
XX
PI Aguera M, Belin MF, Byk T, Honnorat J, Kolattukudy P;
Quach TT, Sobel A;
XX
DR WPI; 1998-449610/39.
DR N-PSDB; AAW60818.

PS Claim 1; Fig 9; 90pp; French.

XX This sequence represents the mouse Unc-33-like phospho-protein (ULIP)-1.
CC The coding sequence was isolated based on similarity to the rat ULIP
CC sequence. Proteins of the ULIP family or their corresponding nucleic
CC acids can be used in compositions for treating neurodegenerative
CC disorders and neoplasms, especially for para-neoplastic neurological
CC syndromes and/or for the early diagnosis of tumourigenesis.

XX Sequence 572 AA;

Query Match 5.4%; Score 31; DB 19; Length 572;

Best Local Similarity 100.0%; Pred. No. 1.2e-21;

Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 276 GEPITASLGTDGSHYWSKNWAKAAAFVTSPP 306

Db 276 gepitaslgtgdshywsknwakaafvtspp 306

RESULT 4

AAW68487
ID AAW68487 standard; Protein; 572 AA.

AC AAW68487;

DT 08-DEC-1998 (first entry)

DE Mouse ULIP-3 protein.

XX Mouse; Unc-33-like phospho-protein; ULIP; rat; neoplasm; tumourigenesis;
KW neurodegenerative disorder; diagnosis.

XX Mus musculus.

XX FR2759701-A1.

XX 21-AUG-1998.

XX 19-FEB-1997; 97FR-0001961.

XX 19-FEB-1997; 97FR-0001961.

XX (INRM) INSERM INST NAT SANTE & RECH MEDICALE.

XX Aguera M, Belin MF, Byk T, Honnorat J, Kolattukudy P;

PI Quach TT, Sobel A;

XX WPI; 1998-449610/39.

DR N-PSDB; AAV60816..

XX Mouse and human ULIP poly:peptide(s) - useful in detection of

PT para-neoplastic neurological syndromes

XX Claim 1; Fig 10; 90pp; French.

XX This sequence represents the mouse Unc-33-like phospho-protein (ULIP)-3.
CC The coding sequence was isolated based on similarity to the rat ULIP
CC sequence. Proteins of the ULIP family or their corresponding nucleic
CC acids can be used in compositions for treating neurodegenerative
CC disorders and neoplasms, especially for para-neoplastic neurological
CC syndromes and/or for the early diagnosis of tumourigenesis.

XX Sequence 572 AA;

Query Match 3.1%; Score 18; DB 19; Length 572;

Best Local Similarity 100.0%; Pred. No. 6.2e-09;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 289 HYWSKNWAKAAAFVTSPP 306

|||||

Db 289 hywsknwakaafvtspp 306

RESULT 5

AAW58948

ID AAW58948 standard; Protein; 167 AA.

XX AAW58948;

XX 27-MAR-2001 (first entry)

XX Breast and ovarian cancer associated antigen protein sequence SEQ ID 656.

XX Human; breast cancer; ovarian cancer; cytostatic; immunosuppressive;
KW neurotropic; neuroprotective; antiviral; antiallergic; hepatotropic;
KW antidiabetic; antiinflammatory; antitumor; vulnary; anticonvulsant;
KW antibacterial; antifungal; antiparasitic; cardiac; immune disorder;
KW Addison's disease; allergy; autoimmune haemolytic anaemia;
KW autoimmune thyroiditis; diabetes mellitus; Crohn's disease;
KW multiple sclerosis; rheumatoid arthritis; ulcerative colitis;
KW cardiovascular disorder; wound healing; neurological disease.

XX Homo sapiens.

XX WO200055173-A1.

XX 21-SEP-2000.

XX 08-MAR-2000; 2000WO-US05881.

XX 12-MAR-1999; 99US-0124270.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Ruben SM;

XX WPI; 2000-611515/58.

XX N-PSDB; AAF21851.

XX New human breast and ovarian cancer associated gene sequences and the
PT polypeptides encoded by these genes, useful in the prevention,
PT treatment and diagnosis of cancer, immune disorders, cardiovascular
PT disorders and neurological diseases -

XX Claim 11; Page 1102; 1299pp; English.

XX Sequences AAF21614 - AAF22031 represent DNA sequences encoding human
CC proteins AAW58711 - AAW59128. The DNA and protein sequences are
CC associated with breast and ovarian cancer. Included in the invention are
CC sequences AAF22032 - AAF22040 and AAW59129 which are used in the
CC isolation and characterisation of the DNA and protein sequences of the
CC invention. The breast and ovarian cancer associated DNA, protein, agonist
CC or antagonist sequences exhibit cytostatic; immunosuppressive;
CC neurotropic; neuroprotective; antiviral; antiallergic; hepatotropic;
CC antidiabetic; antiinflammatory; antitumor; vulnary; anticonvulsant;
CC antibacterial; antifungal; antiparasitic and cardiac activity. The
CC polynucleotide and protein sequences are used in the diagnosis of cancer,
CC particularly breast and ovarian cancer. The nucleic acid sequences,
CC proteins, agonists and antagonists may also be used in the diagnosis,
CC prevention and treatment of immune disorders e.g. Addison's disease,
CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
CC arthritis and ulcerative colitis; cardiovascular disorders such as
CC myocardial ischaemias; wound healing; neurological diseases such as
CC cerebral anoxia and epilepsy; and infectious diseases.

XX Sequence 167 AA;

Query Match 2.6%; Score 15; DB 21; Length 167;

Best Local Similarity 100.0%; Pred. No. 1.8e-06;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 529 PPVRNLHQSGFSLSG 543
 Db 124 ppvrnlhqsgfslsg 138

RESULT 6

AAW54409
 ID AAB54409 standard; Protein; 40 AA.

XX AC AAB54409;

XX DT 09-MAR-2001 (first entry)

XX Human pancreatic cancer antigen protein sequence SEQ ID NO:861.

XX Human; pancreas; pancreatic cancer; pancreatic cancer antigen;
 KW detection; diagnosis; identification; cytostatic; neuroprotective;
 KW neoplastic; immunomodulatory; relaxant; contraceptive; gynaecological;
 KW antiinflammatory; cardiac; gene therapy; chromosome mapping;
 KW linkage analysis; tissue identification; tissue typing; forensic;
 KW neural; immune system; muscular; reproductive; gastrointestinal;
 KW pulmonary; cardiovascular; renal; proliferative.

XX OS Homo sapiens.

XX WO200055320-A1.

XX PD 21-SEP-2000.

XX PF 08-MAR-2000; 2000WO-US05989.

XX PR 12-MAR-1999; 99US-0124270.

XX PA (HUMA-) HUMAN GENOME SCI INC.

XX PI Rosen CA, Ruben SM;

XX WPI; 2000-579444/54.

XX N-PSDB; AAC99174.

PT New nucleic acid that is a pancreatic cancer antigen for preventing,
 PT treating, or ameliorating a medical condition, particular pancreatic
 PT cancer, or for use in assays for diagnosing a pathological condition -
 XX Claim 11; Page 1320; 1379pp; English.

XX AAC98773 to AAC99231 encode the human pancreatic cancer associated
 CC proteins, called pancreatic cancer antigens, given in AAB54008 to
 CC AAB54466. The human pancreatic cancer antigens have cytostatic,
 CC neuroprotective, neoplastic, immunomodulatory, relaxant, contraceptive,
 CC gynaecological, cardiac and antiinflammatory activities, and can be used
 CC in gene therapy. The polynucleotide and proteins can be used for
 CC preventing, treating, or ameliorating a medical condition or in assays
 CC for diagnosing a pathological condition or a susceptibility to one in a
 CC subject. Binding partners to the proteins and the activity of the
 CC proteins can be identified. The pancreatic cancer antigens can be used to
 CC detect, treat or prevent pancreatic disorders, especially cancer.
 CC Agonists and antagonists to the antigens can be screened for. The
 CC pancreatic cancer antigen polynucleotides can be used to design nucleic
 CC acid hybridisation probes that can be used in chromosome mapping, linkage
 CC analysis, tissue identification and/or typing and a variety of forensic
 CC and diagnostic methods. The proteins can be used to generate antibodies
 CC which are used to purify, detect and target the polypeptides, including
 CC both in vivo and in vitro diagnostic and therapeutic methods. The
 CC proteins can be used to treat or prevent neural, immune system, muscular,
 CC reproductive, gastrointestinal, pulmonary, cardiovascular, renal or
 CC proliferative disorders. AAC99232 to AAC99240 and AAB54467 represent
 CC sequences used in the exemplification of the present invention.

XX SQ Sequence 40 AA;

Query Match 2.3%; Score 13; DB 21; Length 40;

Best Local Similarity 100.0%; Pred. No. 4.8e-05;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 91 QGTRKAALAGGTTM 103

Db 26 qgtkaalaggttm 38

RESULT 7

AAW69259
 ID AAW69259 standard; peptide; 27 AA.

XX AC AAW69259;

XX DT 29-OCT-1998 (first entry)

XX DMS protein fragment #1.

XX DMS protein fragment; dense microsphere; DMS; cerebral amyloid formation;
 KW antibody recognition sequence; mammalian brain; inhibitor; therapy;
 KW Alzheimer's disease; senile amyloid plaque.

XX OS Homo sapiens.

XX WO9834643-A1.

XX PD 13-AUG-1998.

XX PF 06-FEB-1998; 98WO-CA00065.

XX PR 03-FEB-1998; 98US-0017689.

XX PR 07-FEB-1997; 97US-0038694.

XX PA (NYMO-) NYMOX CORP.

XX PI Averbach P;

XX WPI; 1998-446954/38.

XX Medicaments preventing the formation of cerebral amyloid of the
 PT brain - by inhibiting synthesis, growth and disruption of dense
 PT microspheres, useful for preventing or treating Alzheimer's disease
 XX Claim 3; Page 23; 43pp; English.

XX This sequence represents a fragment of the dense microsphere (DMS)
 CC protein. This sequence is recognised by an antibody used in the
 CC medicament used in the method of the invention. The method is for
 CC preventing the formation of cerebral amyloid in the mammalian brain, and
 CC comprises administering a medicament that prevents or inhibits the
 CC synthesis, growth and/or disruption of DMS. The method is used for the
 CC prevention, treatment or the prophylaxis of Alzheimer's disease by
 CC impeding the formation of senile amyloid plaques.

XX SQ Sequence 27 AA;

Query Match

Best Local Similarity 2.1%; Score 12; DB 19; Length 27;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 215 LGITGPEGHVLVS 226

Db 4 lgitgpeghvlvs 15

RESULT 8

AAW69260
 ID AAW69260 standard; peptide; 18 AA.

XX AC AAW69260;

XX DT 29-OCT-1998 (first entry)

XX DMS protein fragment #2.
DE DMS protein fragment; dense microsphere; DMS; cerebral amyloid formation;
KW antibody recognition sequence; mammalian brain; inhibitor; therapy;
KW Alzheimer's disease; senile amyloid plaque.
XX Homo sapiens.
XX WO9834643-A1.
XX 13-AUG-1998.
XX 06-FEB-1998; 98WO-CA00065.
XX 03-FEB-1998; 98US-0017689.
XX 07-FEB-1997; 97US-0038694.
XX (NYMO-) NYMOX CORP.
XX Averbach P;
XX WPI; 1998-446954/38.
XX Medicaments preventing the formation of cerebral amyloid of the
PT brain - by inhibiting synthesis, growth and disruption of dense
PT microspheres, useful for preventing or treating Alzheimer's disease
XX Claim 3; Page 23; 43pp; English.
XX This sequence represents a fragment of the dense microsphere (DMS)
CC protein. This sequence is recognised by an antibody used in the
CC medicament used in the method of the invention. The method is for
CC preventing the formation of cerebral amyloid in the mammalian brain, and
CC comprises administering a medicament that prevents or inhibits the
CC synthesis, growth and/or disruption of DMS. The method is used for the
CC prevention, treatment or the prophylaxis of Alzheimer's disease by
CC impeding the formation of senile amyloid plaques.
XX Sequence 18 AA;
XX
XX Query Match 1.9%; Score 11; DB 19; Length 18;
XX Best Local Similarity 100.0%; Pred. No. 0.0022;
XX Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 402 AVGSDADLVW 412
XX | | | | | | | | | |
XX Db 2 avgdadlvw 12
XX
XX RESULT 9
XX AAW69262
XX ID AAW69262 standard; peptide; 20 AA.
XX AC AAW69262;
XX 29-OCT-1998 (first entry)
XX DE DMS protein fragment #4.
XX DMS protein fragment; dense microsphere; DMS; cerebral amyloid formation;
KW antibody recognition sequence; mammalian brain; inhibitor; therapy;
KW Alzheimer's disease; senile amyloid plaque.
XX Homo sapiens.
XX WO9834643-A1.
XX 13-AUG-1998.
XX 06-FEB-1998; 98WO-CA00065.
XX

PR 03-FEB-1998; 98US-0017689.
PR 07-FEB-1997; 97US-0038694.
XX (NYMO-) NYMOX CORP.
XX Averbach P;
XX WPI; 1998-446954/38.
XX Medicaments preventing the formation of cerebral amyloid of the
PT brain - by inhibiting synthesis, growth and disruption of dense
PT microspheres, useful for preventing or treating Alzheimer's disease
XX Claim 3; Page 23; 43pp; English.
XX This sequence represents a fragment of the dense microsphere (DMS)
CC protein. This sequence is recognised by an antibody used in the
CC medicament used in the method of the invention. The method is for
CC preventing the formation of cerebral amyloid in the mammalian brain, and
CC comprises administering a medicament that prevents or inhibits the
CC synthesis, growth and/or disruption of DMS. The method is used for the
CC prevention, treatment or the prophylaxis of Alzheimer's disease by
CC impeding the formation of senile amyloid plaques.
XX Sequence 20 AA;
XX
XX Query Match 1.9%; Score 11; DB 19; Length 20;
XX Best Local Similarity 100.0%; Pred. No. 0.0024;
XX Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 24 IVNDQSFYAD 34
XX | | | | | | | | | |
XX Db 1 ivndqsfyad 11
XX
XX RESULT 10
XX AAR69874
XX ID AAR69874 standard; Protein; 460 AA.
XX AC AAR69874;
XX 27-OCT-1995 (first entry)
XX DE B.thermoglucodasius D-hydantoinase with mutated C-terminal region.
XX D-hydantoinase; dihydropyrimidinase; mutein; variant; mutation;
KW D-N-carbamoyl-alpha-amino acid; penicillin; cephalosporin;
KW production; racemisation.
XX OS Bacillus thermoglucodasius.
XX PN DE4328829-A.
XX 02-MAR-1995.
XX 27-AUG-1993; 93DE-4328829.
XX 27-AUG-1993; 93DE-4328829.
XX (BOEF) BOEHRINGER MANNHEIM GMBH.
XX Burtcher H, Lang G, Popp F;
XX WPI; 1995-099331/14.
XX N-PSDB; AAQ80565.
XX New highly active and stable D-hydantoinase enzyme - isolated
PT from Bacillus thermoglucodasius, useful in the prodn. of
PT D-N-carbamoyl-alpha-amino acid for synthesis of, e.g. penicillin
XX Claim 1; Page 5-6; 8pp; German.
XX

CC The wild-type D-hydantoinase gene was isolated from *Bacillus*
 CC *thermoglucodasius* using the amplification primers Hyd1 and Hyd2 (see
 CC AAQ80566 and AAQ80567). A single-base deletion in the HindIII site (from
 CC AAGCTT to AAGCT) was then introduced to give a coding region (AAQ80565)
 CC that encoded a D-hydantoinase (AAR69874) that was shorter than the
 CC wild-type protein with a different C-terminal sequence (see Features
 CC Table). The mutant enzyme was found to have improved heat stability
 CC and higher activity compared to known hydantoinases.
 XX
 SQ Sequence 460 AA;

Query Match 1.7%; Score 10; DB 16; Length 460;
 Best Local Similarity 100.0%; Pred. No. 0.34;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 402 AVGSADLVI 411
 |||||
 Db 386 avgsadlvi 395

RESULT 11

AAW03025
 ID AAW03025 standard; Protein; 485 AA.

AC AAW03025;

DT 31-JAN-1997 (first entry)

DE Pseudomonas hydantoinase.

KW Hydantoinase; hydantoin; d-N-carbamoyl-alpha-amino acid;
 KW intermediate; drug synthesis; penicillin; cephalosporin;
 KW antibiotic.

OS Pseudomonas sp.

XX WO9620275-A1.

XX 04-JUL-1996.

XX 26-DEC-1995; 95WO-JP02688.

XX 28-DEC-1994; 94JP-0326865.

XX (KANF) KANEGAFUCHI KAGAKU KOGYO KK.

PI Ikenaka Y, Nanba H, Takahashi S, Takano M, Yajima K;
 PI Yamada Y;

DR WPI; 1996-321848/32.

DR N-PSDB; AAT31258.

PT Prodn. of D-N-carbamoyl amino acid from 5-substd. hydantoin - using
 PT a recombinant hydantoinase derived from a strain of *Pseudomonas*,
 PT *Agrobacterium* or *Bacillus*

PS Claim 19; Page 36-40; 54pp; English.

CC D-N-carbamoyl-alpha-amino acid is produced from a 5-substituted
 CC hydantoin by treatment with a hydantoinase expressed by a
 CC transformant microorganism carrying a vector containing DNA coding
 CC for the hydantoinase and derived from *Bacillus* sp. KKK245,
 CC *Agrobacterium* sp. KKK712 or *Pseudomonas* sp. KKK003A. The D-N-
 CC carbamoyl-alpha-amino acid can be used for the production of
 CC optically active amino acids (especially D-phenylglycine and D-
 CC p-hydroxyphenylglycine) as intermediates for drug synthesis,
 CC especially for the production of semi-synthetic penicillin and
 CC cephalosporin antibiotics.

XX Sequence 485 AA;

Query Match 1.7%; Score 10; DB 17; Length 485;
 Best Local Similarity 100.0%; Pred. No. 0.35;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 378 NEFVAVTSTN 387
 |||||
 Db 358 nefvavtstn 367

RESULT 12

AAG46570

ID AAG46570 standard; Protein; 576 AA.

AC AAG46570;

DT 18-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 58602.

KW Protein identification; signal transduction pathway; metabolic pathway;
 KW hybridisation assay; genetic mapping; gene expression control; promoter;
 KW termination sequence.

OS Arabidopsis thaliana.

XX EP1033405-A2.

XX 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

XX 05-MAR-1999; 99US-0123180.

XX 09-MAR-1999; 99US-0123548.

XX 23-MAR-1999; 99US-0125788.

XX 25-MAR-1999; 99US-0126264.

XX 29-MAR-1999; 99US-0126785.

XX 01-APR-1999; 99US-0127462.

XX 06-APR-1999; 99US-0128234.

XX 08-APR-1999; 99US-0128714.

XX 16-APR-1999; 99US-0129845.

XX 19-APR-1999; 99US-0130077.

XX 21-APR-1999; 99US-0130449.

XX 23-APR-1999; 99US-0130510.

XX 23-APR-1999; 99US-0130891.

XX 28-APR-1999; 99US-0131449.

XX 30-APR-1999; 99US-0132048.

XX 30-APR-1999; 99US-0132407.

XX 04-MAY-1999; 99US-0132484.

XX 05-MAY-1999; 99US-0132485.

XX 06-MAY-1999; 99US-0132486.

XX 07-MAY-1999; 99US-0132487.

XX 07-MAY-1999; 99US-0132863.

XX 11-MAY-1999; 99US-0134256.

XX 14-MAY-1999; 99US-0134218.

XX 14-MAY-1999; 99US-0134219.

XX 14-MAY-1999; 99US-0134221.

XX 14-MAY-1999; 99US-0134370.

XX 18-MAY-1999; 99US-0134768.

XX 19-MAY-1999; 99US-0134941.

XX 20-MAY-1999; 99US-0135124.

XX 21-MAY-1999; 99US-0135353.

XX 24-MAY-1999; 99US-0135629.

XX 25-MAY-1999; 99US-0136021.

XX 27-MAY-1999; 99US-0136392.

XX 28-MAY-1999; 99US-0136782.

XX 01-JUN-1999; 99US-0137222.

XX 03-JUN-1999; 99US-0137528.

XX 04-JUN-1999; 99US-0137502.

XX 07-JUN-1999; 99US-0137724.

XX 08-JUN-1999; 99US-0138094.

XX 10-JUN-1999; 99US-0138540.

XX 10-JUN-1999; 99US-0138847.

PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139452.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.

PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151080.
PR 31-AUG-1999; 99US-0151303.
PR 01-SEP-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 1.7%; Score 10; DB 21; Length 576;

Best Loca. Similarity 100.0%; Pred. No. 0.41; Mismatches 0; Indels 0; Gaps 0;

Matches 10; Conservative 0;

Qy 214 ELGITGPEGH 223

Db 135 e.gitgpegh 144

RESULT 13
AAG46569
ID AAG46569 standard; Protein; 588 AA.
XX AAG46569;
AC AAG46569;
DT 18-OCT-2000 (first entry)
XX Arabidopsis thaliana protein fragment SEQ ID NO: 58601.
DE Arabidopsis thaliana
XX Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX Arabidopsis thaliana.
OS Arabidopsis thaliana.
XX EPI033405-A2.
XX 06-SEP-2000.
XX 25-FEB-2000; 2000EP-0301439.
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 18-MAY-1999; 99US-0134370.
PR 19-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145919.
PR 02-AUG-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.

PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 22-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 1.7%; Score 10; DB 21; Length 588;
Best Local Similarity 100.0%; Pred. No. 0.42;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 214 ELGITGPEGH 223
Db 147 elgitgpegh 156

RESULT 14
AAG46568
ID AAG46568 standard; Protein; 683 AA.
XX
AC AAG46568;

XX 18-OCT-2000 (first entry)
DE Arabidopsis thaliana protein fragment SEQ ID NO: 58600.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX Arabidopsis thaliana.
OS
XX
PN EP1033405-A2.
XX
PD
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.

```
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148555.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149388.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.

PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 1.7%; Score 10; DB 21; Length 683;
Best Local Similarity 100.0%; Pred. No. 0.47;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 214 ELGHTGPEGH 223
Db 242 elgtgpegh 251

RESULT 15
AAG23636
ID AAG23636 standard; Protein; 283 AA.
XX
AC AAG23636;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 27018.
XX Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
```

KW termination sequence.

XX OS Arabidopsis thaliana.
XX PN EP1033405-A2.
XX PD 06-SEP-2000.
XX PF 25-FEB-2000; 2000EP-0301439.
XX PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126284.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 07-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135224.
PR 24-MAY-1999; 99US-0135353.
PR 25-MAY-1999; 99US-0135629.
PR 27-MAY-1999; 99US-0136021.
PR 28-MAY-1999; 99US-0136392.
PR 01-JUN-1999; 99US-0136782.
PR 03-JUN-1999; 99US-0137222.
PR 04-JUN-1999; 99US-0137528.
PR 07-JUN-1999; 99US-0137502.
PR 08-JUN-1999; 99US-0137724.
PR 10-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138840.
PR 14-JUN-1999; 99US-0138847.
PR 16-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 18-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151338.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.

PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158023.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 1.6%; Score 9; DB 21; Length 283;

Best Local Similarity 100.0%; Pred. No. 2.1;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 95 AALAGGTTM 103

Db 125 aalagggtm 133

|||||||

RESULT 16

AAG23635

ID AAG23635 standard; Protein; 284 AA.

XX AC AAG23635;

XX DT 17-OCT-2000 (first entry)

XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 27017.

XX KW Protein identification; signal transduction pathway; metabolic pathway;

XX KW hybridisation assay; genetic mapping; gene expression control; promoter;

XX KW termination sequence.

XX OS Arabidopsis thaliana.

XX PN EP1033405-A2.

XX PD 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.

XX WO200058473-A2.
 PN
 XX
 PD
 XX
 PF 31-MAR-2000; 2000WO-US08621.
 XX
 PR 31-MAR-1999; 99US-0127607.
 PR 02-APR-1999; 99US-0127636.
 PR 05-APR-1999; 99US-0127728.
 PR 30-MAR-2000; 2000US-0540763.
 XX
 PA (CURA-) CURAGEN CORP.
 XX
 PI Shimkets RA, Leach M;
 XX
 XX
 DR WPI; 2000-602362/57.
 DR N-PSDB; AAC77230.
 XX

Novel nucleic acids and peptides derived from open reading frame X,
 useful for treating e.g. cancers, proliferative disorders,
 neurodegenerative disorders and cardiovascular disease -

PS Claim 11; Page 4756-4757; 5507pp; English.

AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
 which represent the human ORFX open reading frames 1 to 3161. The ORFX
 sequences have activities such as: cytostatic; hepatotropic; vulnery;
 antiparasitic; antiparkinsonian; nootropic; neuroprotective;
 osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
 immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
 antidiabetic; hypotensive; dermatological; immunosuppressive;
 antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
 antithyroid; and antianaemic. The sequences can be used for determining
 the presence of or predisposition to, or preventing or treating
 pathological conditions associated with an ORFX-associated disorder. The
 nucleic acids can be used to express ORFX proteins in gene therapy
 vectors. The proteins and nucleic acids may be used to treat cancers,
 proliferative disorders, neurodegenerative disorders, osteoarthritis,
 graft vs host disease, cardiovascular disease, diabetes mellitus,
 hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
 erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
 bacterial or fungal infection, malaria, autoimmune disorders, asthma,
 allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
 nocturnal haemoglobinuria, antiinflammatory disease; to enhance
 coagulation; to inhibit thrombosis; and as a contraceptive.

XX Sequence 169 AA;

Query Match 1.4%; Score 8; DB 21; Length 169;
 Best Local Similarity 100.0%; Pred. No. 13;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 429 VEYNIFEG 436
 |||||
 Db 21 veynifeg 28

RESULT 18
 AAR82837
 ID AAR82837 standard; Protein; 457 AA.

XX AAR82837;

DT 13-MAR-1996 (first entry)

DE Agrobacterium radiobacter D-hydantoinase.

XX D-hydantoinase; D-N-carbamylase; enzyme; stereospecific reaction;
 KW D-amino acid.
 XX

OS Agrobacterium radiobacter.

XX EP677585-A1.
 PN
 XX
 PD 18-OCT-1995.
 XX
 PF 24-MAR-1995; 95EP-0104393.
 XX
 PR 15-APR-1994; 94IT-0M10726.
 XX
 PA (ENIE) ENIRICERCH SPA.
 XX
 PI Frascotti G, Galli G, Grandi G, Grifantini R;
 XX
 DR WPI; 1995-352764/46.
 DR N-PSDB; AAT01498.
 XX

Prod'n. of D-alpha amino acids from racemic 5-substd. hydantoin cpds.
 PT - using microorganisms contg. hydantoinase and carbamylase genes.

PS Disclosure; Fig.5A-5C; 44pp; English.

XX A. radiobacter is the donor microorganism for genes encoding
 CC D-hydantoinase and D-N-carbamylase which are expressed in
 CC Escherichia coli or Bacillus subtilis using plasmid pSM651. The
 CC resulting recombinant E. coli may be used to catalyze the
 CC stereospecific preparation of D-amino acids from racemic
 CC 5-substituted hydantoin compounds.

XX Sequence 457 AA;

Query Match 1.4%; Score 8; DB 16; Length 457;
 Best Local Similarity 100.0%; Pred. No. 30;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 401 VAVGSDAD 408
 |||||
 Db 382 vavgsdad 389

RESULT 19

AAR60325
 ID AAR60325 standard; Protein; 526 AA.

XX AAR60325;

AC 01-MAR-1995 (first entry)

DE Streptomyces cholesterol oxidase.

XX Cholesterol oxidase; Streptomyces; enzyme; thermostability;
 KW diagnosis; endocrine disease; metabolism.

OS Streptomyces sp. SA-COO.

XX JP06169765-A.

PD 21-JUN-1994.

XX 05-JUL-1993; 93JP-0165558.

XX 05-OCT-1992; 92JP-0266130.

XX (TOYM) TOYOBO KK.

XX WPI; 1994-237584/29.

XX N-PSDB; AAQ70418.

XX Cholesterol oxidase with high thermostability - useful for the
 PT diagnosis of endocrine diseases

XX Claim 4; Page 11-13; 15pp; Japanese.

XX

CC Cholesterol oxidase is used for measurement of cholesterol which is
 CC an index for the diagnosis of endocrine diseases and abnormal
 CC metabolism. The enzyme is obtained by adding amino acid residues to
 CC the N-terminus of cholesterol oxidase derived from Streptomyces sp.
 CC SA-COO, which shows the same activity as that of the naturally
 CC occurring enzyme, but has higher thermostability. Cholesterol
 CC oxidase DNA sequences derived from Streptomyces sp. are given in
 CC AAQ70417-18.
 XX
 SQ Sequence 526 AA;

Query Match 1.4%; Score 8; DB 15; Length 526;
 Best Local Similarity 100.0%; Pred. No. 34;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 95 AALAGGTT 102
 |||||
 Db 1 aalaggtt 8

RESULT 20
 AAW08932
 ID AAW08932 standard; Protein; 546 AA.

XX AC AAW08932;

XX DT 18-FEB-1997 (first entry)

XX DE Cholesterol oxidase, Val121Ala.

XX KW Cholesterol oxidase; mutation; thermal stability; substitution.

XX OS Streptomyces sp. SA-COO.

XX FH Key Location/Qualifiers
 FT Misc-difference 121
 FT /label= Val121Ala

XX PN JP08242860-A.

XX PD 24-SEP-1996.

XX PF 07-MAR-1995; 95JP-0047339.

XX PR 07-MAR-1995; 95JP-0047339.

XX PA (TOYM) TOYOB0 KK.

XX DR WPI; 1996-479904/48.

XX PT Modified cholesterol oxidase enzyme - having higher thermal
 PT stability than the parent enzyme, useful in cholesterol determin.

XX PS Claim 9; Page 11-12; 23pp; Japanese.

XX CC The sequences given in AAW08931-36 represent modified cholesterol
 CC oxidases. These modified cholesterol oxidases may be prepared by
 CC mutation of the wild type sequence (see also AAW08930). These mutant
 CC cholesterol oxidases pref. have amino acid substitutions at positions
 CC 103Ser, 121Val, 135Arg and/or 145Val They have higher thermal stability
 CC than the parent enzyme.

XX SQ Sequence 546 AA;

Query Match 1.4%; Score 8; DB 17; Length 546;
 Best Local Similarity 100.0%; Pred. No. 35;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 95 AALAGGTT 102
 |||||

Db 21 aalaggtt 28

RESULT 21
 AAW08933

ID AAW08933 standard; Protein; 546 AA.

XX AC AAW08933;

XX DT 18-FEB-1997 (first entry)

XX DE Cholesterol oxidase, Arg135His.

XX KW Cholesterol oxidase; mutation; thermal stability; substitution.

XX OS Streptomyces sp. SA-COO.

XX FH Key Location/Qualifiers
 FT Misc-difference 135
 FT /label= Arg135His

XX PN JP08242860-A.

XX PD 24-SEP-1996.

XX PF 07-MAR-1995; 95JP-0047339.

XX PR 07-MAR-1995; 95JP-0047339.

XX PA (TOYM) TOYOB0 KK.

XX DR WPI; 1996-479904/48.

XX PT Modified cholesterol oxidase enzyme - having higher thermal
 PT stability than the parent enzyme, useful in cholesterol determin.

XX PS Claim 11; Page 13-14; 23pp; Japanese.

XX CC The sequences given in AAW08931-36 represent modified cholesterol
 CC oxidases. These modified cholesterol oxidases may be prepared by
 CC mutation of the wild type sequence (see also AAW08930). These mutant
 CC cholesterol oxidases pref. have amino acid substitutions at positions
 CC 103Ser, 121Val, 135Arg and/or 145Val They have higher thermal stability
 CC than the parent enzyme.

XX SQ Sequence 546 AA;

Query Match 1.4%; Score 8; DB 17; Length 546;
 Best Local Similarity 100.0%; Pred. No. 35;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 95 AALAGGTT 102
 |||||
 Db 21 aalaggtt 28

RESULT 22
 AAW08934

ID AAW08934 standard; Protein; 546 AA.

XX AC AAW08934;

XX DT 18-FEB-1997 (first entry)

XX DE Cholesterol oxidase, Val145Glu.

XX KW Cholesterol oxidase; mutation; thermal stability; substitution.

XX OS Streptomyces sp. SA-COO.

XX FH Key Location/Qualifiers

FT Misc-difference 145
XX /label= Vall145Glu
PN JP08242860-A.
XX
XX PD 24-SEP-1996.
XX PF 07-MAR-1995; 95JP-0047339.
XX PR 07-MAR-1995; 95JP-0047339.
XX PA (TOYM) TOYOBOKK.
XX WPI; 1996-479904/48.
XX
XX Modified cholesterol oxidase enzyme - having higher thermal
PT stability than the parent enzyme, useful in cholesterol determin.
PT reagent
XX
XX Claim 13; Page 14-15; 23pp; Japanese.
XX
XX The sequences given in AAW08931-36 represent modified cholesterol
CC oxidases. These modified cholesterol oxidases may be prepared by
CC mutation of the wild type sequence (see also AAW08930). These mutant
CC cholesterol oxidases pref. have amino acid substitutions at positions
CC 103Ser, 121Val, 135Arg and/or 145Val They have higher thermal stability
CC than the parent enzyme.
XX
XX SQ Sequence 546 AA;

Query Match 1.4%; Score 8; DB 17; Length 546;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 95 AALAGGTT 102
DB 21 aalagttt 28
|||||

RESULT 23
AAW08935
ID AAW08935 standard; Protein; 546 AA.
XX
XX AC AAW08935;
XX
XX DT 18-FEB-1997 (first entry)
XX
XX DE Cholesterol oxidase, Ser103Thr, Vall145Glu.
XX
XX KW Cholesterol oxidase; mutation; thermal stability; substitution.
XX
XX OS Streptomyces sp. SA-COO.
XX
XX PH Key Location/Qualifiers
FT Misc-difference 103
FT /label= Ser103Thr
FT Misc-difference 121
FT /label= Vall121Ala
FT Misc-difference 145
FT /label= Vall145Glu
XX
XX PN JP08242860-A.
XX
XX PD 24-SEP-1996.
XX
XX PF 07-MAR-1995; 95JP-0047339.
XX
XX PR 07-MAR-1995; 95JP-0047339.
XX
XX PA (TOYM) TOYOBOKK.
XX
XX DR WPI; 1996-479904/48.
XX
XX Modified cholesterol oxidase enzyme - having higher thermal
PT stability than the parent enzyme, useful in cholesterol determin.
PT reagent
XX
XX Claim 17; Page 17-18; 23pp; Japanese.
XX
XX The sequences given in AAW08931-36 represent modified cholesterol
CC oxidases. These modified cholesterol oxidases may be prepared by
CC mutation of the wild type sequence (see also AAW08930). These mutant
CC cholesterol oxidases pref. have amino acid substitutions at positions
CC 103Ser, 121Val, 135Arg and/or 145Val They have higher thermal stability
CC than the parent enzyme.
XX
XX SQ Sequence 546 AA;

PT stability than the parent enzyme, useful in cholesterol determin.
PT reagent
XX
XX Claim 15; Page 16-17; 23pp; Japanese.
XX
XX The sequences given in AAW08931-36 represent modified cholesterol
CC oxidases. These modified cholesterol oxidases may be prepared by
CC mutation of the wild type sequence (see also AAW08930). These mutant
CC cholesterol oxidases pref. have amino acid substitutions at positions
CC 103Ser, 121Val, 135Arg and/or 145Val They have higher thermal stability
CC than the parent enzyme.
XX
XX SQ Sequence 546 AA;

Query Match 1.4%; Score 8; DB 17; Length 546;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 95 AALAGGTT 102
DB 21 aalagttt 28
|||||

RESULT 24
AAW08936
ID AAW08936 standard; Protein; 546 AA.
XX
XX AC AAW08936;
XX
XX DT 18-FEB-1997 (first entry)
XX
XX DE Cholesterol oxidase, Ser103Thr, Vall121Ala, Vall145Glu.
XX
XX KW Cholesterol oxidase; mutation; thermal stability; substitution.
XX
XX OS Streptomyces sp. SA-COO.
XX
XX PH Key Location/Qualifiers
FT Misc-difference 103
FT /label= Ser103Thr
FT Misc-difference 121
FT /label= Vall121Ala
FT Misc-difference 145
FT /label= Vall145Glu
XX
XX PN JP08242860-A.
XX
XX PD 24-SEP-1996.
XX
XX PF 07-MAR-1995; 95JP-0047339.
XX
XX PR 07-MAR-1995; 95JP-0047339.
XX
XX PA (TOYM) TOYOBOKK.
XX
XX DR WPI; 1996-479904/48.
XX
XX Modified cholesterol oxidase enzyme - having higher thermal
PT stability than the parent enzyme, useful in cholesterol determin.
PT reagent
XX
XX Claim 17; Page 17-18; 23pp; Japanese.
XX
XX The sequences given in AAW08931-36 represent modified cholesterol
CC oxidases. These modified cholesterol oxidases may be prepared by
CC mutation of the wild type sequence (see also AAW08930). These mutant
CC cholesterol oxidases pref. have amino acid substitutions at positions
CC 103Ser, 121Val, 135Arg and/or 145Val They have higher thermal stability
CC than the parent enzyme.
XX
XX SQ Sequence 546 AA;

Query Match 1.4%; Score 8; DB 17; Length 546;
 Best Local Similarity 100.0%; Pred. No. 35;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 95 AALAGGTT 102
 |||||
 Db 21 aalaggtt 28

RESULT 25

AAW08930
 ID AAW08930 standard; Protein; 546 AA.

XX AC AAW08930;

XX DT 18-FEB-1997 (first entry)

XX DE Wild type cholesterol oxidase.

XX KW Cholesterol oxidase; mutation; thermal stability; substitution.

XX OS Streptomyces sp. SA-COO.

XX PN JP08242860-A.

XX PD 24-SEP-1996.

XX PF 07-MAR-1995; 95JP-0047339.

XX PR 07-MAR-1995; 95JP-0047339.

XX PA (TOYM) TOYOBO KK.

XX DR WPI; 1996-479904/48.

XX DR N-PSDB; AAT49320.

XX PT Modified cholesterol oxidase enzyme - having higher thermal
 stability than the parent enzyme, useful in cholesterol determin.

XX PT reagent

XX PS Claim 3; Page 8-9; 23pp; Japanese.

XX CC This sequence represents wild type cholesterol oxidase. Modified
 cholesterol oxidases may be prepared by mutation of this
 sequence. Mutant cholesterol oxidases pref. have amino acid
 substitutions at positions 103Ser, 121Val, 135Arg and/or 145Val (see
 also AAW08931-36). The modified enzymes have higher thermal stability
 than the parent enzyme.

XX SQ Sequence 546 AA;

Query Match 1.4%; Score 8; DB 17; Length 546;
 Best Local Similarity 100.0%; Pred. No. 35;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 95 AALAGGTT 102
 |||||
 Db 21 aalaggtt 28

RESULT 26

AAW08931
 ID AAW08931 standard; Protein; 546 AA.

XX AC AAW08931;

XX DT 18-FEB-1997 (first entry)

XX DE Cholesterol oxidase, Ser103Thr.

XX KW Cholesterol oxidase; mutation; thermal stability; substitution.

XX OS Streptomyces sp. SA-COO.
 XX FH Key Location/Qualifiers
 FT Misc-difference 103
 XX FT /label= Ser103Thr

XX PN JP08242860-A.

XX PD 24-SEP-1996.

XX PF 07-MAR-1995; 95JP-0047339.

XX PR 07-MAR-1995; 95JP-0047339.

XX PA (TOYM) TOYOBO KK.

XX DR WPI; 1996-479904/48.

XX PT Modified cholesterol oxidase enzyme - having higher thermal
 stability than the parent enzyme, useful in cholesterol determin.

XX PT reagent

XX PS Claim 7; Page 10-11; 23pp; Japanese.

XX CC The sequences given in AAW08931-36 represent modified cholesterol
 oxidases. These modified cholesterol oxidases may be prepared by
 mutation of the wild type sequence (see also AAW08930). These mutant
 cholesterol oxidases pref. have amino acid substitutions at positions
 103Ser, 121Val, 135Arg and/or 145Val They have higher thermal stability
 than the parent enzyme.

XX SQ Sequence 546 AA;

Query Match 1.4%; Score 8; DB 17; Length 546;
 Best Local Similarity 100.0%; Pred. No. 35;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 95 AALAGGTT 102
 |||||
 Db 21 aalaggtt 28

RESULT 27

AAW75425

ID AAW75425 standard; Protein; 837 AA.

XX AC AAW75425;

XX DT 30-MAR-1999 (first entry)

XX DE Human aggrecan degrading metalloprotease 1.

XX KW Human; aggrecan degrading metalloprotease; cartilage; proteoglycan;
 interglobular domain; matrix metalloprotease; bovine; interleukin-1beta;
 primer; PCR; amplification; inhibitor; cleavage; inhibitor; ADMP;
 arthritis; joint injury; pseudogout.

XX OS Homo sapiens.

XX PN WO9905291-A2.

XX PD 04-FEB-1999.

XX PF 24-JUL-1998; 98WO-US15438.

XX PR 16-OCT-1997; 97US-0062169.

XX PR 25-JUL-1997; 97US-0053850.

XX PR 15-AUG-1997; 97US-0055836.

XX PA (DUPO) DU PONT PHARM CO.

XX XX

PI Arner EC, Burn TC, Copeland RA, Decicco CP, Liu R, Magolda R;
 PI Pratta M, Solomon KA, Tortorella MD, Trzaskos JM, Yang F;
 DR WPI; 1999-142943/12.
 DR N-PSDB; AAX00725.
 XX
 XX
 PT New isolated aggrecan degrading metalloproteases - used to develop
 PT products for treating, e.g. Osteoarthritis, joint injury, reactive
 PT arthritis, psoriatic arthritis or juvenile rheumatoid arthritis
 XX
 XX
 PS Claim 25; Page 61-62; 73pp; English.
 XX
 CC This sequence represents the human aggrecan degrading metalloprotease 1
 CC (ADMP-1). ADMP-1 and ADMP-2 (AAW75426) are novel proteases that cleave
 CC the aggrecan (a major cartilage proteoglycan) between residues
 CC Glu373-Ala374 of the interglobular domain (compared with cleavage between
 CC Asn341-Phe342 by the matrix metalloproteases MMP-1, -2, -3, -7, -8, -9
 CC and -13). ADMP-1 and ADMP-2 were isolated and purified from the
 CC conditioned media of bovine nasal cartilage stimulated by
 CC interleukin-1beta. The purified proteins were partially sequenced and
 CC primers were synthesised based on the resultant amino acid sequences
 CC (AAX00727-X00732 for ADMP-1 and AAX00733-X00736 for ADMP-2). These were
 CC used to isolate the corresponding genes from cDNA. The ADMP polypeptides
 CC can be used for identifying inhibitors of ADMP activity which would
 CC prevent cleavage of the aggrecan core protein, thereby decreasing the
 CC loss of aggrecan from cartilage. Such inhibitors can be used for treating
 CC diseases such as osteoarthritis, joint injury, reactive arthritis, acute
 CC pyrophosphate arthritis (pseudogout), psoriatic arthritis and juvenile
 CC rheumatoid arthritis.
 XX
 SQ Sequence 837 AA;

Query Match 1.4%; Score 8; DB 20; Length 837;
 Best Local Similarity 100.0%; Pred. No. 51;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 512 PGSGAPAR 519
 Db 73 pgsgapar 80
 |||||
 |||||

RESULT 28
 AAY99429
 ID AAY99429 standard; Protein; 837 AA.
 XX
 AC AAY99429;
 XX
 DT 08-AUG-2000 (first entry)
 XX
 DE Human PRO1563 (UNQ769) amino acid sequence SEQ ID NO:317.
 XX
 DE Human; PRO polypeptide; membrane bound protein; receptor; diagnosis;
 KW transmembrane; secretion; immunoadhesion; pharmaceutical; screening.
 XX
 KW Homo sapiens.
 OS
 XX WO200012708-A2.
 PN
 XX
 PD 09-MAR-2000.
 XX
 PF 01-SEP-1999; 99WO-US20111.
 XX
 XX 01-SEP-1998; 98US-0098716.
 PR 01-SEP-1998; 98US-0098749.
 PR 01-SEP-1998; 98US-0098750.
 PR 02-SEP-1998; 98US-0098803.
 PR 02-SEP-1998; 98US-0098821.
 PR 02-SEP-1998; 98US-0098843.
 PR 09-SEP-1998; 98US-0099536.
 PR 09-SEP-1998; 98US-0099596.
 PR 09-SEP-1998; 98US-0099598.
 PR 09-SEP-1998; 98US-0099602.

PR 09-SEP-1998; 98US-0099642.
 PR 10-SEP-1998; 98US-0099741.
 PR 10-SEP-1998; 98US-0099754.
 PR 10-SEP-1998; 98US-0099763.
 PR 10-SEP-1998; 98US-0099792.
 PR 10-SEP-1998; 98US-0099808.
 PR 10-SEP-1998; 98US-0099812.
 PR 10-SEP-1998; 98US-0099815.
 PR 10-SEP-1998; 98US-0099816.
 PR 15-SEP-1998; 98US-0100385.
 PR 15-SEP-1998; 98US-0100388.
 PR 15-SEP-1998; 98US-0100390.
 PR 16-SEP-1998; 98US-0100584.
 PR 16-SEP-1998; 98US-0100627.
 PR 16-SEP-1998; 98US-0100661.
 PR 16-SEP-1998; 98US-0100662.
 PR 16-SEP-1998; 98US-0100664.
 PR 17-SEP-1998; 98US-0100683.
 PR 17-SEP-1998; 98US-0100684.
 PR 17-SEP-1998; 98US-0100710.
 PR 17-SEP-1998; 98US-0100711.
 PR 17-SEP-1998; 98US-0100919.
 PR 18-SEP-1998; 98US-0100930.
 PR 18-SEP-1998; 98US-0100848.
 PR 18-SEP-1998; 98US-0100849.
 PR 18-SEP-1998; 98US-0101014.
 PR 18-SEP-1998; 98US-0101068.
 PR 18-SEP-1998; 98US-0101071.
 PR 22-SEP-1998; 98US-0101279.
 PR 23-SEP-1998; 98US-0101471.
 PR 23-SEP-1998; 98US-0101472.
 PR 23-SEP-1998; 98US-0101474.
 PR 23-SEP-1998; 98US-0101475.
 PR 23-SEP-1998; 98US-0101476.
 PR 23-SEP-1998; 98US-0101477.
 PR 23-SEP-1998; 98US-0101479.
 PR 24-SEP-1998; 98US-0101738.
 PR 24-SEP-1998; 98US-0101741.
 PR 24-SEP-1998; 98US-0101743.
 PR 24-SEP-1998; 98US-0101915.
 PR 24-SEP-1998; 98US-0101916.
 PR 29-SEP-1998; 98US-0102207.
 PR 29-SEP-1998; 98US-0102240.
 PR 29-SEP-1998; 98US-0102307.
 PR 29-SEP-1998; 98US-0102330.
 PR 29-SEP-1998; 98US-0102331.
 PR 30-SEP-1998; 98US-0102484.
 PR 30-SEP-1998; 98US-0102487.
 PR 30-SEP-1998; 98US-0102570.
 PR 30-SEP-1998; 98US-0102571.
 PR 01-OCT-1998; 98US-0102684.
 PR 01-OCT-1998; 98US-0102687.
 PR 02-OCT-1998; 98US-0102965.
 PR 06-OCT-1998; 98US-0103258.
 PR 06-OCT-1998; 98US-0103449.
 PR 07-OCT-1998; 98US-0103314.
 PR 07-OCT-1998; 98US-0103315.
 PR 07-OCT-1998; 98US-0103328.
 PR 07-OCT-1998; 98US-0103395.
 PR 07-OCT-1998; 98US-0103396.
 PR 07-OCT-1998; 98US-0103401.
 PR 08-OCT-1998; 98US-0103633.
 PR 08-OCT-1998; 98US-0103678.
 PR 08-OCT-1998; 98US-0103679.
 PR 08-OCT-1998; 98US-0103711.
 PR 14-OCT-1998; 98US-0104257.
 PR 20-OCT-1998; 98US-0104987.
 PR 20-OCT-1998; 98US-0105000.
 PR 20-OCT-1998; 98US-0105002.
 PR 21-OCT-1998; 98US-0105104.
 PR 22-OCT-1998; 98US-0105169.
 PR 22-OCT-1998; 98US-0105266.
 PR 26-OCT-1998; 98US-0105693.

```

PR 26-OCT-1998; 98US-0105694.
PR 27-OCT-1998; 98US-0105807.
PR 27-OCT-1998; 98US-0105881.
PR 27-OCT-1998; 98US-0105882.
PR 27-OCT-1998; 98US-0106062.
PR 28-OCT-1998; 98US-0106023.
PR 28-OCT-1998; 98US-0106029.
PR 28-OCT-1998; 98US-0106030.
PR 28-OCT-1998; 98US-0106032.
PR 28-OCT-1998; 98US-0106033.
PR 28-OCT-1998; 98US-0106178.
PR 29-OCT-1998; 98US-0106248.
PR 29-OCT-1998; 98US-0106384.
PR 29-OCT-1998; 98US-0108500.
PR 30-OCT-1998; 98US-0106464.
PR 03-NOV-1998; 98US-0106856.
PR 03-NOV-1998; 98US-0106902.
PR 03-NOV-1998; 98US-0106905.
PR 03-NOV-1998; 98US-0106919.
PR 03-NOV-1998; 98US-0106932.
PR 03-NOV-1998; 98US-0106934.
PR 10-NOV-1998; 98US-0107783.
PR 17-NOV-1998; 98US-0108775.
PR 17-NOV-1998; 98US-0108787.
PR 17-NOV-1998; 98US-0108788.
PR 17-NOV-1998; 98US-0108801.
PR 17-NOV-1998; 98US-0108802.
PR 17-NOV-1998; 98US-0108806.
PR 17-NOV-1998; 98US-0108807.
PR 17-NOV-1998; 98US-0108867.
PR 17-NOV-1998; 98US-0108925.
PR 18-NOV-1998; 98US-0108848.
PR 18-NOV-1998; 98US-0108849.
PR 18-NOV-1998; 98US-0108850.
PR 18-NOV-1998; 98US-0108851.
PR 18-NOV-1998; 98US-0108852.
PR 18-NOV-1998; 98US-0108858.
PR 18-NOV-1998; 98US-0108904.
XX
XX (GETH ) GENENTECH INC.
XX Baker K, Goddard A, Gurney AL, Smith V, Watanabe CK, Wood WI;
XX WPI; 2000-237871/20.
XX N-PSDB; AAA37111.
XX
XX New mammalian DNA sequences encoding transmembrane, receptor or
XX secreted PRO polypeptides, useful for screening of potential peptide or
XX small molecule inhibitors of the relevant receptor/ligand interactions
XX
XX Claim 12; Fig 180; 773pp; English.
XX
XX AAA37022 to AAA37144 encode the new isolated human transmembrane,
XX receptor or secreted PRO polypeptides given in AAY99340 to AAY99462. The
XX transmembrane and receptor PRO proteins can be used for screening of
XX potential peptide or small molecule inhibitors of the relevant
XX receptor/ligand interactions. The polypeptides and nucleotide sequences
XX encoding them have various industrial applications, including uses as
XX pharmaceutical and diagnostic agents. AAA37145 to AAA37330 represent
XX PCR primers and hybridization probes used in the isolation of the PRO
XX polypeptides from the present invention.
XX
XX Sequence 837 AA;
XX
XX Query Match 1.4%; Score 8; DB 21; Length 837;
XX Best Local Similarity 100.0%; Pred. No. 51;
XX Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 512 PGSGAPAR 519
XX |||||||
XX 73 pgsgapar 80
XX
XX RESULT 30
XX AAW10617
XX ID AAW10617 standard; Protein; 15 AA.
XX
XX AC AAW106.7;

```

```

RESULT 29
AAB66178
ID AAB66173 standard; protein; 837 AA.
XX
XX AC AAB66173;
XX
XX 02-APR-2001 (first entry)
XX
XX Protein of the invention #90.
XX
XX Secreted; transmembrane; gene therapy.
XX
XX Unidentified.
XX
XX WO200078961-A1.
XX
XX 28-DEC-2000.
XX
XX 18-FEB-2000; 2000WO-US04342.
XX
XX 23-JUN-1999; 99US-0141037.
XX
XX 26-JUL-1999; 99US-0144758.
XX
XX 01-SEP-1999; 99WO-US20111.
XX
XX 29-OCT-1999; 99US-0162506.
XX
XX 30-NOV-1999; 99WO-US28313.
XX
XX 02-DEC-1999; 99WO-US28551.
XX
XX 16-DEC-1999; 99WO-US30095.
XX
XX 05-JAN-2000; 2000WO-US00219.
XX
XX 06-JAN-2000; 2000WO-US00376.
XX
XX (GETH ) GENENTECH INC.
XX
XX Baker KP, Botstein D, Desnoyers L, Eaton DL, Ferrara N, Fong S;
XX Gao W, Goddard A, Godowski PJ, Grimaldi CJ, Gurney AL, Hillan KJ;
XX Pan J, Paoni NF, Roy MA, Smith V, Stewart TA, Tumas D;
XX Watanabe CK, Williams PM, Wood WI;
XX WPI; 2001-071395/08.
XX
XX Secreted and transmembrane proteins and nucleic acids designated PRO,
XX useful as hybridization probes, in chromosome and gene mapping and gene
XX therapy -
XX
XX Claim 1; Fig 180; 787pp; English.
XX
XX The present invention relates to secreted and transmembrane proteins.
XX These proteins and the DNA encoding them may be used as hybridization
XX probes, in chromosome and gene mapping and in the generation of
XX anti-sense RNA and DNA. They may also be used to generate either
XX transgenic animals or knockout animals which are in turn useful for
XX development and screening of therapeutically useful reagents.
XX The nucleic acids may also be used in gene therapy.
XX
XX Sequence 837 AA;
XX
XX Query Match 1.4%; Score 8; DB 22; Length 837;
XX Best Local Similarity 100.0%; Pred. No. 51;
XX Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 512 PGSGAPAR 519
XX |||||||
XX 73 pgsgapar 80
XX
XX RESULT 30
XX AAW10617
XX ID AAW10617 standard; Protein; 15 AA.
XX
XX AC AAW106.7;

```

XX DT 29-OCT-1997 (first entry)
 XX DE L-galactono-1,4-lactone:ferricytochrome-c-oxidoreductase fragment 5.
 XX KW L-galactono-gamma-lactone; ascorbic acid; cauliflower; mitochondrion;
 XX KW antioxidant; chiral; L-galactonate; D-galacturonic acid.
 XX OS Brassica oleracea botrytis.
 XX PN WO9704100-A2.
 XX PD 06-FEB-1997.
 XX PF 17-JUL-1996; 96WO-DE01339.
 XX PR 17-JUL-1995; 95DE-4025991.
 XX PA (FRAU) FRAUNHOFER GES FOERDERUNG ANGEWANDTEN.
 XX PI Wissler J;
 XX DR WPI; 1997-132640/12.
 XX DR N-PSDB; AAT61001.
 XX PT Enzyme for oxidn. of L-galactono-gamma-lactone to ascorbic acid -
 XX PT useful for ascorbic acid synthesis from waste materials and determn.
 XX PT of chiral L-galactonate or D-galacturonic acid
 XX PS Claim 5; Page 26; 34pp; German.
 XX CC A new L-galactono-1,4-lactone:ferricytochrome-c-oxidoreductase
 CC enzyme was isolated from cauliflower mitochondria. The enzyme
 CC oxidises L-galactono-gamma-lactone to ascorbic acid, has an
 CC apparent molecular weight of 56 kD and includes the amino acid
 CC sequence: Val-Gln-Gln-Leu-Val-Asp-Ala-Ile-Gln-Glu-Tyr-Gly-Leu.
 CC A further 15 partial fragments of the enzyme have been obtained
 CC and sequenced, including the present sequence. The enzyme is used
 CC for ascorbic acid synthesis, particularly from waste materials,
 CC and for the determination of chiral L-galactonate or D-galacturonic
 CC acid. Ascorbic acid is well known as a nutritional supplement, free
 CC radical scavenger and food antioxidant.
 XX SQ Sequence 15 AA;

Query Match 1.2%; Score 7; DB 18; Length 15;
 Best Local Similarity 100.0%; Pred. No. 15;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 149 KEELEAL 155
 Db 7 keeleal 13
 |||||

RESULT 31
 AAW69340
 ID AAW69340 standard; Protein; 55 AA.
 XX AC AAW69340;
 XX DT 25-NOV-1998 (first entry)
 XX DE Secreted protein of clone CD311_2.
 XX KW Secreted protein; nutritional activity; immune stimulant;
 KW immune suppressor; haematopoiesis regulator; tissue growth activity;
 KW activin/inhibin activity; chemotactic/chemokinetic activity;
 KW haemostatic activity; thrombolytic activity; receptor/ligand activity;
 KW anti-inflammatory activity; cadherin suppressor; tumour inhibitor;
 KW tumour invasion suppressor; therapy.
 XX OS Homo sapiens.

XX PN WO9837094-A2.
 XX PD 27-AUG-1998.
 XX PF 24-FEB-1998; 98WO-US03595.
 XX PR 23-FEB-1998; 98US-0028168.
 XX PR 24-FEB-1997; 97US-0804561.
 XX PA (GEMY) GENETICS INST INC.
 XX PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D;
 PI Racie LA, Spaulding V, Treacy M;
 XX DR WPI; 1998-467491/40.
 XX DR N-PSDB; AAV58536.
 XX PT Secreted proteins with biological activity - and encoding
 PT polynucleotide(s), useful e.g. in therapy and diagnosis of medical
 PT conditions and to identify protein agonists or antagonists
 XX PS Claim 15; Page 69; 111pp; English.
 XX CC This sequence is a secreted protein of the invention. This sequence
 CC is encoded by the DNA of clone CD311_2, which was isolated from a human
 CC foetal brain cDNA library. The polypeptides are predicted to have
 CC useful biological activities which would make them suitable for treating,
 CC preventing or ameliorating medical conditions in humans and animals,
 CC although no supporting biological data is given. Suggested activities
 CC include nutritional activity, immune stimulating or suppressing activity,
 CC haematopoiesis regulating activity, tissue growth activity,
 CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC activity, cadherin/tumour invasion suppressor activity, anti-inflammatory
 CC activity or other activities. They may also be used for diagnostic
 CC purposes. The polynucleotides are useful to produce the proteins or
 CC generate probes or primers to identify and/or amplify similar genes
 CC e.g. species homologues. They are also useful for gene therapy and to
 CC produce transgenic animals with altered gene expression.
 XX SQ Sequence 55 AA;

Query Match 1.2%; Score 7; DB 19; Length 55;
 Best Local Similarity 100.0%; Pred. No. 46;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 539 FSLSGSQ 545
 Db 43 fslsgsq 49
 |||||

RESULT 32
 AAG33074
 ID AAG33074 standard; Protein; 90 AA.
 XX AC AAG33074;
 XX DT 18-OCT-2000 (first entry)
 XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 40020.
 XX KW Protein identification; signal transduction pathway; metabolic pathway;
 KW hybridisation assay; genetic mapping; gene expression control; promoter;
 KW termination sequence.
 XX OS Arabidopsis thaliana.
 XX PN EP1033405-A2.
 XX PD 06-SEP-2000.

PF XX 25-FEB-1999; 2000EP-0301439.
25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0123788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0128845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 24-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 22-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 23-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 02-AUG-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147410.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.

```
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 1.2%; Score 7; DB 21; Length 90;
Best Local Similarity 100.0%; Pred. No. 71;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 328 VTGSANC 334
Db 30 vtgsahc 36

RESULT 33
AAG48928
ID AAG48928 standard; Protein; 90 AA.
XX AC AAG48928;
XX DT 18-OCT-2000 (first entry)
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 61841.
XX DX
XX KW Protein identification; signal transduction pathway; metabolic pathway;
XX KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX KW termination sequence.
XX OS Arabidopsis thaliana.
XX PN EP1033405-A2.
XX PD 06-SEP-2000.
XX PF 25-FEB-2000; 2000EP-0301439.
XX PR 25-FEB-1999; 99US-0121825.
XX PR 05-MAR-1999; 99US-0123180.
XX PR 09-MAR-1999; 99US-0123548.
XX PR 23-MAR-1999; 99US-0125788.
XX PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134250.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
```

```
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147030.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147410.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154779.
PR 20-SEP-1999; 99US-0155039.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.

PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match. 1.2%; Score 7; DB 21; Length 90;
Best Local Similarity 100.0%; Pred. No. 71;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Caps 0;

Qy 328 VTGSAHC 334
Db 30 vtgsahc 36
|||||

RESULT 34
AAG33073
ID AAG33073 standard; Protein; 91 AA.
XX
AC AAG33073;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 40019.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
```


PR 14-OCT-1999; 99US-0159637.
 PR 14-OCT-1999; 99US-0159638.
 PR 18-OCT-1999; 99US-0159584.
 PR 21-OCT-1999; 99US-0160741.
 PR 21-OCT-1999; 99US-0160767.
 PR 21-OCT-1999; 99US-0160768.
 PR 21-OCT-1999; 99US-0160770.
 PR 21-OCT-1999; 99US-0160814.
 PR 21-OCT-1999; 99US-0160815.
 PR 22-OCT-1999; 99US-0160980.
 PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160989.
 PR 23-OCT-1999; 99US-0161404.
 PR 23-OCT-1999; 99US-0161405.
 PR 25-OCT-1999; 99US-0161406.
 PR 26-OCT-1999; 99US-0161359.
 PR 26-OCT-1999; 99US-0161360.
 PR 26-OCT-1999; 99US-0161361.
 PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161992.
 PR 28-OCT-1999; 99US-0161993.
 PR 29-OCT-1999; 99US-0162142.
 PR 29-OCT-1999; 99US-0162142.

Query Match 1.2% Score 7; DB 21; Length 91;
 Best Local Similarity 100.0%; Pred. No. 71;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 328 VTGSAHC 334
 Db 31 vtgsahc 37

RESULT 35

AAG48927
 ID AAG48927 standard; Protein; 91 AA.

XX AC AAG48927;

XX DT 18-OCT-2000 (first entry)

XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 61840.

KW Protein identification; signal transduction pathway; metabolic pathway;
 KW hybridisation assay; genetic mapping; gene expression control; promoter;
 KW termination sequence.

XX OS Arabidopsis thaliana.

XX PN EP1033405-A2.

XX PD 06-SEP-2000.

XX PF 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.
 PR 05-MAR-1999; 99US-0123180.
 PR 09-MAR-1999; 99US-0123548.
 PR 23-MAR-1999; 99US-0125788.
 PR 23-MAR-1999; 99US-0126264.
 PR 29-MAR-1999; 99US-0126785.
 PR 01-APR-1999; 99US-0127462.
 PR 06-APR-1999; 99US-0128234.
 PR 08-APR-1999; 99US-0128714.
 PR 16-APR-1999; 99US-0129845.
 PR 19-APR-1999; 99US-0130077.
 PR 21-APR-1999; 99US-0130449.
 PR 23-APR-1999; 99US-0130510.
 PR 23-APR-1999; 99US-0130891.
 PR 28-APR-1999; 99US-0131449.
 PR 30-APR-1999; 99US-0132048.
 PR 30-APR-1999; 99US-0132407.
 PR 04-MAY-1999; 99US-0132484.
 PR 05-MAY-1999; 99US-0132485.

PR 06-MAY-1999; 99US-0132486.
 PR 06-MAY-1999; 99US-0132487.
 PR 07-MAY-1999; 99US-0132863.
 PR 11-MAY-1999; 99US-0134256.
 PR 14-MAY-1999; 99US-0134218.
 PR 14-MAY-1999; 99US-0134219.
 PR 14-MAY-1999; 99US-0134221.
 PR 14-MAY-1999; 99US-0134370.
 PR 18-MAY-1999; 99US-0134768.
 PR 19-MAY-1999; 99US-0134941.
 PR 20-MAY-1999; 99US-0135124.
 PR 21-MAY-1999; 99US-0135353.
 PR 24-MAY-1999; 99US-0135629.
 PR 25-MAY-1999; 99US-0136021.
 PR 27-MAY-1999; 99US-0136392.
 PR 28-MAY-1999; 99US-0136782.
 PR 01-JUN-1999; 99US-0137222.
 PR 03-JUN-1999; 99US-0137528.
 PR 04-JUN-1999; 99US-0137502.
 PR 07-JUN-1999; 99US-0137724.
 PR 08-JUN-1999; 99US-0138094.
 PR 10-JUN-1999; 99US-0138540.
 PR 10-JUN-1999; 99US-0138847.
 PR 14-JUN-1999; 99US-0139119.
 PR 16-JUN-1999; 99US-0139452.
 PR 16-JUN-1999; 99US-0139453.
 PR 17-JUN-1999; 99US-0139492.
 PR 18-JUN-1999; 99US-0139454.
 PR 18-JUN-1999; 99US-0139455.
 PR 18-JUN-1999; 99US-0139456.
 PR 18-JUN-1999; 99US-0139457.
 PR 18-JUN-1999; 99US-0139458.
 PR 18-JUN-1999; 99US-0139459.
 PR 18-JUN-1999; 99US-0139460.
 PR 18-JUN-1999; 99US-0139461.
 PR 18-JUN-1999; 99US-0139462.
 PR 18-JUN-1999; 99US-0139463.
 PR 18-JUN-1999; 99US-0139750.
 PR 18-JUN-1999; 99US-0139763.
 PR 21-JUN-1999; 99US-0139817.
 PR 22-JUN-1999; 99US-0139899.
 PR 23-JUN-1999; 99US-0140353.
 PR 23-JUN-1999; 99US-0140354.
 PR 24-JUN-1999; 99US-0140695.
 PR 28-JUN-1999; 99US-0140823.
 PR 29-JUN-1999; 99US-0140991.
 PR 30-JUN-1999; 99US-0141287.
 PR 01-JUL-1999; 99US-0141842.
 PR 01-JUL-1999; 99US-0142154.
 PR 02-JUL-1999; 99US-0142055.
 PR 06-JUL-1999; 99US-0142390.
 PR 08-JUL-1999; 99US-0142803.
 PR 09-JUL-1999; 99US-0142920.
 PR 12-JUL-1999; 99US-0142977.
 PR 13-JUL-1999; 99US-0143542.
 PR 14-JUL-1999; 99US-0143624.
 PR 15-JUL-1999; 99US-0144005.
 PR 16-JUL-1999; 99US-0144085.
 PR 16-JUL-1999; 99US-0144086.
 PR 19-JUL-1999; 99US-0144325.
 PR 19-JUL-1999; 99US-0144331.
 PR 19-JUL-1999; 99US-0144332.
 PR 19-JUL-1999; 99US-0144333.
 PR 19-JUL-1999; 99US-0144334.
 PR 19-JUL-1999; 99US-0144335.
 PR 20-JUL-1999; 99US-0144352.
 PR 20-JUL-1999; 99US-0144632.
 PR 20-JUL-1999; 99US-0144884.
 PR 21-JUL-1999; 99US-0144814.
 PR 21-JUL-1999; 99US-0145086.
 PR 21-JUL-1999; 99US-0145088.
 PR 22-JUL-1999; 99US-0145085.
 PR 22-JUL-1999; 99US-0145087.

PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 06-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159395.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 18-OCT-1999; 99US-0159638.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.

PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161993.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 1.2%; Score 7; DB 21; Length 91;
Best Local Similarity 100.0%; Pred. No. 71;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 328 VTGSAHC 334
Db 31 vtgsahc 37
|||||

RESULT 36

AAB80092
ID AAB80092 standard; Protein; 93 AA.

AC AAB80092;

XX 30-APR-2001 (first entry)

DE Corynebacterium glutamicum MP protein sequence SEQ ID NO:918.

XX Corynebacterium glutamicum;
KW fine chemical production; microorganism; organic acid; nucleoside;
KW nonproteinogenic amino acid; purine base; pyrimidine base; nucleotide;
KW lipid; saturated fatty acid; unsaturated fatty acid; diol; vitamin;
KW carbohydrate; aromatic compound; cofactor; polyketide; enzyme.

OS Corynebacterium glutamicum.

XX WO200100843-A2.

PN 04-JAN-2001.

XX 23-JUN-2000; 2000WO-IB00923.

XX 25-JUN-1999; 99US-0141031.

PR 01-JUL-1999; 99DE-1030476.

PR 02-JUL-1999; 99US-0142101.

PR 08-JUL-1999; 99DE-1031415.

PR 08-JUL-1999; 99DE-1031418.

PR 08-JUL-1999; 99DE-1031419.

PR 08-JUL-1999; 99DE-1031420.

PR 08-JUL-1999; 99DE-1031424.

PR 08-JUL-1999; 99DE-1031434.

PR 08-JUL-1999; 99DE-1031435.

PR 08-JUL-1999; 99DE-1031443.

PR 08-JUL-1999; 99DE-1031453.

PR 08-JUL-1999; 99DE-1031457.

PR 08-JUL-1999; 99DE-1031465.

PR 08-JUL-1999; 99DE-1031478.

PR 08-JUL-1999; 99DE-1031510.

PR 08-JUL-1999; 99DE-1031541.

PR 08-JUL-1999; 99DE-1031573.

PR 08-JUL-1999; 99DE-1031592.

PR 08-JUL-1999; 99DE-1031632.

PR 08-JUL-1999; 99DE-1031634.

PR 08-JUL-1999; 99DE-1031636.

PR 09-JUL-1999; 99DE-1032125.

PR 09-JUL-1999; 99DE-1032126.
 PR 09-JUL-1999; 99DE-1032130.
 PR 09-JUL-1999; 99DE-1032186.
 PR 09-JUL-1999; 99DE-1032206.
 PR 09-JUL-1999; 99DE-1032227.
 PR 09-JUL-1999; 99DE-1032228.
 PR 09-JUL-1999; 99DE-1032229.
 PR 09-JUL-1999; 99DE-1032230.
 PR 14-JUL-1999; 99DE-1032922.
 PR 14-JUL-1999; 99DE-1032926.
 PR 14-JUL-1999; 99DE-1032928.
 PR 14-JUL-1999; 99DE-1033004.
 PR 14-JUL-1999; 99DE-1033005.
 PR 14-JUL-1999; 99DE-1033006.
 PR 12-AUG-1999; 99US-0148613.
 PR 27-AUG-1999; 99DE-1040764.
 PR 27-AUG-1999; 99DE-1040765.
 PR 27-AUG-1999; 99DE-1040766.
 PR 27-AUG-1999; 99DE-1040832.
 PR 31-AUG-1999; 99DE-1041378.
 PR 31-AUG-1999; 99DE-1041379.
 PR 31-AUG-1999; 99DE-1041380.
 PR 31-AUG-1999; 99DE-1041394.
 PR 03-SEP-1999; 99DE-1041396.
 PR 03-SEP-1999; 99DE-1042076.
 PR 03-SEP-1999; 99DE-1042077.
 PR 03-SEP-1999; 99DE-1042079.
 PR 03-SEP-1999; 99DE-1042086.
 PR 03-SEP-1999; 99DE-1042087.
 PR 03-SEP-1999; 99DE-1042088.
 PR 03-SEP-1999; 99DE-1042095.
 PR 03-SEP-1999; 99DE-1042124.
 PR 03-SEP-1999; 99DE-1042129.
 PR 09-MAR-2000; 2000US-0187970.
 XX (BADI) BASF AG.

XX Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G;
 WPI: 2001-137957/14.
 DR N-PSDB; AAF72211.
 XX Nucleic acids from Corynebacterium glutamicum encoding metabolic
 PT pathway proteins, useful for producing fine chemicals in
 PT microorganisms, including organic acids, nonproteinogenic amino acids,
 PT and purine and pyrimidine bases -
 XX

PS Claim 20; Page 1416; 1737pp; English.

XX AAF71753 to AAF72330 encode the Corynebacterium glutamicum metabolic
 CC pathway (MP) proteins given in AAB79634 to AAB80211. The C. glutamicum
 CC MP nucleic acids are useful for the production of fine chemicals
 CC in microorganisms, including organic acids, nonproteinogenic amino
 CC acids, purine and pyrimidine bases, nucleosides, nucleotides, lipids,
 CC saturated and unsaturated fatty acids, diols, carbohydrates, aromatic
 CC compounds, vitamins, cofactors, polyketides and enzymes.
 XX

SQ Sequence 93 AA;

Query Match 1.2%; Score 7; DB 22; Length 93;
 Best Local Similarity 100.0%; Pred. No. 73;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 316 HLTCLLS 322

Db 10 hltclls 16
 |||||

RESULT 37

AAW63722
 XX AAW63722 standard; Protein; 110 AA.

AC AAW63722;
 XX
 DT 08-OCT-1998 (first entry)
 XX
 DE C. histolyticum CHCI protein N-terminal.
 XX
 KW Class I collagenase; CHCI; isolation; transplant; immunotherapy; ulcer;
 KW identification; gene therapy; tumour.
 XX
 OS Clostridium histolyticum.
 XX
 PN WO9822574-A2.
 XX
 XX 28-MAY-1998.
 XX
 PF 30-OCT-1997; 97WO-EP06002.
 XX
 XX 24-JAN-1997; 97EP-0101085.
 PR 19-NOV-1996; 96EP-0118490.
 XX
 PA (BOEF) BOEHRINGER MANNHEIM GMBH.
 XX
 PI Ambrosius D, Burtischer H, Hesse F;
 XX
 DR WPI; 1998-312460/27.
 XX
 XX New recombinant collagenase type I from Clostridium histolyticum -
 CC useful for isolating cells from animal tissue, e.g. for
 CC transplantation or immuno-therapy and in gene therapy
 PT
 PS Claim 1; Page 27; 38pp; German.
 XX
 CC This sequence represents the N-terminal from a Class I collagenase, CHCI
 CC from Clostridium histolyticum. This protein can be used to isolate cells
 CC from animal or human tissue, e.g. for use in transplantation or
 CC immuno-therapy (particularly pancreatic islets). This protein may also be
 CC used for identification of rare cells carrying specific markers, for use
 CC in gene therapy of a wide variety of tissues, including solid tumours,
 CC and for cleaning ulcers. Recombinant CHCI has high enzymatic activity and
 CC is very stable. It can be produced very pure (free of other clostridial
 CC enzymes and toxins) and in high yield. Compositions containing the
 CC protein may be used where cells are to be recovered without alteration of
 CC their cell-surface antigens.
 XX
 SQ Sequence 110 AA;

Query Match 1.2%; Score 7; DB 19; Length 110;
 Best Local Similarity 100.0%; Pred. No. 84;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 114 VSLAAY 120

Db 28 vsllaay 34
 |||||

RESULT 38

AAG24039
 ID AAG24039 standard; Protein; 121 AA.

XX
 AC AAG24039;

XX
 DT 17-OCT-2000 (first entry)

XX
 DE Arabidopsis thaliana protein fragment SEQ ID NO: 27563.

XX
 KW Protein identification; signal transduction pathway; metabolic pathway;
 KW hybridisation assay; genetic mapping; gene expression control; promoter;
 KW termination sequence.

XX Arabidopsis thaliana.

OS Arabidopsis thaliana.

XX EP1033405-A2.

PN

XX PD 06-SEP-2000. 99US-0142055.
XX PF 25-FEB-2000; 2000EP-0301439. 99US-0142390.
XX PF 25-FEB-1999; 99US-0142803.
XX PF 05-MAR-1999; 99US-0142920.
XX PF 09-MAR-1999; 99US-0142977.
XX PF 23-MAR-1999; 99US-0143542.
XX PF 25-MAR-1999; 99US-0143624.
XX PF 29-MAR-1999; 99US-0144005.
XX PF 01-APR-1999; 99US-0144085.
XX PF 06-APR-1999; 99US-0144086.
XX PF 08-APR-1999; 99US-0144325.
XX PF 16-APR-1999; 99US-0144331.
XX PF 19-APR-1999; 99US-0144332.
XX PF 21-APR-1999; 99US-0144333.
XX PF 23-APR-1999; 99US-0144334.
XX PF 28-APR-1999; 99US-0144335.
XX PF 30-APR-1999; 99US-0144352.
XX PF 30-APR-1999; 99US-0144632.
XX PF 04-MAY-1999; 99US-0144884.
XX PF 05-MAY-1999; 99US-0145086.
XX PF 06-MAY-1999; 99US-0145088.
XX PF 07-MAY-1999; 99US-0145085.
XX PF 11-MAY-1999; 99US-0145087.
XX PF 14-MAY-1999; 99US-0145089.
XX PF 14-MAY-1999; 99US-0145192.
XX PF 14-MAY-1999; 99US-0145145.
XX PF 14-MAY-1999; 99US-0145218.
XX PF 14-MAY-1999; 99US-0145224.
XX PF 18-MAY-1999; 99US-0145276.
XX PF 19-MAY-1999; 99US-0145913.
XX PF 20-MAY-1999; 99US-0145918.
XX PF 21-MAY-1999; 99US-0145919.
XX PF 24-MAY-1999; 99US-0145951.
XX PF 25-MAY-1999; 99US-0146386.
XX PF 27-MAY-1999; 99US-0146388.
XX PF 01-JUN-1999; 99US-0146389.
XX PF 03-JUN-1999; 99US-0147038.
XX PF 04-JUN-1999; 99US-0147204.
XX PF 07-JUN-1999; 99US-0147302.
XX PF 08-JUN-1999; 99US-0147192.
XX PF 10-JUN-1999; 99US-0147260.
XX PF 10-JUN-1999; 99US-0147303.
XX PF 10-JUN-1999; 99US-0147303.
XX PF 14-JUN-1999; 99US-0147416.
XX PF 16-JUN-1999; 99US-0147935.
XX PF 16-JUN-1999; 99US-0147935.
XX PF 17-JUN-1999; 99US-0148171.
XX PF 18-JUN-1999; 99US-0148319.
XX PF 18-JUN-1999; 99US-0148341.
XX PF 18-JUN-1999; 99US-0148565.
XX PF 18-JUN-1999; 99US-0148684.
XX PF 18-JUN-1999; 99US-0149368.
XX PF 18-JUN-1999; 99US-0149175.
XX PF 18-JUN-1999; 99US-0149426.
XX PF 18-JUN-1999; 99US-0149722.
XX PF 18-JUN-1999; 99US-0149722.
XX PF 18-JUN-1999; 99US-0149902.
XX PF 18-JUN-1999; 99US-0149902.
XX PF 18-JUN-1999; 99US-0149930.
XX PF 18-JUN-1999; 99US-0150566.
XX PF 18-JUN-1999; 99US-0150884.
XX PF 18-JUN-1999; 99US-0151065.
XX PF 18-JUN-1999; 99US-0151065.
XX PF 21-JUN-1999; 99US-0151080.
XX PF 22-JUN-1999; 99US-0151080.
XX PF 23-JUN-1999; 99US-0151303.
XX PF 23-JUN-1999; 99US-0140353.
XX PF 24-JUN-1999; 99US-0140354.
XX PF 24-JUN-1999; 99US-0140695.
XX PF 24-JUN-1999; 99US-0140823.
XX PF 29-JUN-1999; 99US-0140991.
XX PF 30-JUN-1999; 99US-0141287.
XX PF 01-JUL-1999; 99US-0141842.
XX PF 01-JUL-1999; 99US-0142154.
XX PF 02-JUL-1999; 99US-0142055.
XX PF 06-JUL-1999; 99US-0142390.
XX PF 08-JUL-1999; 99US-0142803.
XX PF 09-JUL-1999; 99US-0142920.
XX PF 12-JUL-1999; 99US-0142977.
XX PF 13-JUL-1999; 99US-0143542.
XX PF 14-JUL-1999; 99US-0143624.
XX PF 15-JUL-1999; 99US-0144005.
XX PF 16-JUL-1999; 99US-0144085.
XX PF 16-JUL-1999; 99US-0144086.
XX PF 19-JUL-1999; 99US-0144325.
XX PF 19-JUL-1999; 99US-0144331.
XX PF 19-JUL-1999; 99US-0144332.
XX PF 19-JUL-1999; 99US-0144333.
XX PF 19-JUL-1999; 99US-0144334.
XX PF 19-JUL-1999; 99US-0144335.
XX PF 20-JUL-1999; 99US-0144352.
XX PF 20-JUL-1999; 99US-0144632.
XX PF 20-JUL-1999; 99US-0144884.
XX PF 21-JUL-1999; 99US-0145086.
XX PF 21-JUL-1999; 99US-0145088.
XX PF 22-JUL-1999; 99US-0145085.
XX PF 22-JUL-1999; 99US-0145087.
XX PF 22-JUL-1999; 99US-0145089.
XX PF 22-JUL-1999; 99US-0145192.
XX PF 23-JUL-1999; 99US-0145145.
XX PF 23-JUL-1999; 99US-0145218.
XX PF 26-JUL-1999; 99US-0145224.
XX PF 26-JUL-1999; 99US-0145276.
XX PF 27-JUL-1999; 99US-0145913.
XX PF 27-JUL-1999; 99US-0145918.
XX PF 27-JUL-1999; 99US-0145919.
XX PF 28-JUL-1999; 99US-0145951.
XX PF 02-AUG-1999; 99US-0146386.
XX PF 02-AUG-1999; 99US-0146388.
XX PF 02-AUG-1999; 99US-0146389.
XX PF 03-AUG-1999; 99US-0147038.
XX PF 04-AUG-1999; 99US-0147204.
XX PF 04-AUG-1999; 99US-0147302.
XX PF 05-AUG-1999; 99US-0147192.
XX PF 05-AUG-1999; 99US-0147260.
XX PF 06-AUG-1999; 99US-0147303.
XX PF 06-AUG-1999; 99US-0147416.
XX PF 09-AUG-1999; 99US-0147935.
XX PF 09-AUG-1999; 99US-0147935.
XX PF 10-AUG-1999; 99US-0148171.
XX PF 11-AUG-1999; 99US-0148319.
XX PF 12-AUG-1999; 99US-0148341.
XX PF 13-AUG-1999; 99US-0148565.
XX PF 13-AUG-1999; 99US-0148684.
XX PF 16-AUG-1999; 99US-0149368.
XX PF 17-AUG-1999; 99US-0149175.
XX PF 18-AUG-1999; 99US-0149426.
XX PF 20-AUG-1999; 99US-0149722.
XX PF 20-AUG-1999; 99US-0149722.
XX PF 20-AUG-1999; 99US-0149902.
XX PF 23-AUG-1999; 99US-0149902.
XX PF 23-AUG-1999; 99US-0149930.
XX PF 25-AUG-1999; 99US-0150566.
XX PF 26-AUG-1999; 99US-0150884.
XX PF 27-AUG-1999; 99US-0151065.
XX PF 27-AUG-1999; 99US-0151065.
XX PF 27-AUG-1999; 99US-0151080.
XX PF 30-AUG-1999; 99US-0151080.
XX PF 31-AUG-1999; 99US-0151303.
XX PF 01-SEP-1999; 99US-0151438.
XX PF 01-SEP-1999; 99US-0151930.
XX PF 07-SEP-1999; 99US-0152363.
XX PF 10-SEP-1999; 99US-0153070.
XX PF 13-SEP-1999; 99US-0153070.
XX PF 15-SEP-1999; 99US-0153758.
XX PF 16-SEP-1999; 99US-0154018.
XX PF 16-SEP-1999; 99US-0154039.
XX PF 20-SEP-1999; 99US-0154779.

PR 22-SEP-1999; 99US-0155139.
 PR 23-SEP-1999; 99US-0155486.
 PR 24-SEP-1999; 99US-0155659.
 PR 28-SEP-1999; 99US-0156458.
 PR 29-SEP-1999; 99US-0156596.
 PR 04-OCT-1999; 99US-0157117.
 PR 05-OCT-1999; 99US-0157753.
 PR 06-OCT-1999; 99US-0157865.
 PR 07-OCT-1999; 99US-0158029.
 PR 08-OCT-1999; 99US-0158232.
 PR 12-OCT-1999; 99US-0158369.
 PR 13-OCT-1999; 99US-0159293.
 PR 13-OCT-1999; 99US-0159294.
 PR 13-OCT-1999; 99US-0159295.
 PR 14-OCT-1999; 99US-0159329.
 PR 14-OCT-1999; 99US-0159330.
 PR 14-OCT-1999; 99US-0159331.
 PR 14-OCT-1999; 99US-0159637.
 PR 14-OCT-1999; 99US-0159638.
 PR 18-OCT-1999; 99US-0159584.
 PR 21-OCT-1999; 99US-0160741.
 PR 21-OCT-1999; 99US-0160767.
 PR 21-OCT-1999; 99US-0160768.
 PR 21-OCT-1999; 99US-0160770.
 PR 21-OCT-1999; 99US-0160814.
 PR 21-OCT-1999; 99US-0160815.
 PR 22-OCT-1999; 99US-0160980.
 PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160989.
 PR 25-OCT-1999; 99US-0161404.
 PR 25-OCT-1999; 99US-0161405.
 PR 25-OCT-1999; 99US-0161406.
 PR 26-OCT-1999; 99US-0161359.
 PR 26-OCT-1999; 99US-0161360.
 PR 26-OCT-1999; 99US-0161361.
 PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161992.
 PR 28-OCT-1999; 99US-0161993.
 PR 29-OCT-1999; 99US-0162142.

Query Match 1.2%; Score 7; DB 21; Length 121;

Best Local Similarity 100.0%; Pred. No. 91;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 152 LEALVKE 158

Db 51 lealvke 57

RESULT 39

AAW20410
 ID AAW20410 standard; protein; 124 AA.

AC AAW20410;

DT 29-JUL-1997 (first entry)

DE H. pylori cytoplasmic protein, 3261306.aa.

XX Vaccine; prevention; treatment; infection; identification;
 KW binding compound; bacterium; life cycle; activator; bacteria; inhibitor;
 KW duodenal ulcer disease; chronic gastritis; diagnosis; envelope;
 KW cofactor metabolism.

XX Helicobacter pylori.

XX Key Location/Qualifiers

FT Key

FT Misc-difference 72

FT /label= unknown

FT /note= "encoded by RGC codon"

FT Misc-difference 73

FT /note= "encoded by GTW codon"

XX

PN WO9640893-A1.

XX 19-DEC-1996.

XX 06-JUN-1996; 96WO-US09122.

XX 01-APR-1996; 96US-0630405.

XX 07-JUN-1995; 95US-0487032.

XX (ASTR) ASTRA AB.

XX Berglindh OT, Smith D, Meilgaard BL;

XX WPI; 1997-052306/05.

XX N-PSDB; AAT67792.

XX Helicobacter pylori nucleic acid sequences and related
 PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori
 PT infection, and to detect Helicobacter

XX Claim 61; Page 591; 1481pp; English.

XX The present sequence is a Helicobacter pylori cytoplasmic protein
 CC involved in cofactor metabolism.

CC The protein may be used in a vaccine to prevent or treat H. pylori
 CC infection or to identify H. pylori polypeptide binding compounds,
 CC useful as potential H. pylori life cycle activators or inhibitors.

CC The genomic sequence of H. pylori (ATCC 55679) was determined from
 CC overlapping contigs generated by mechanically shearing the bacterial
 CC DNA. The sequences were analysed for ORF of at least 180 nucleotides,
 CC and the predicted coding regions defined by computer evaluation. To
 CC identify likely H. pylori antigens for vaccine development, the amino
 CC acid sequences predicted from various ORF were analysed for significant
 CC homology to other known or exported membrane proteins. Having identified
 CC and determined the sequences of interest, particular regions can be
 CC isolated from H. pylori by PCR amplification for recombinant polypeptide
 CC production, e.g. in E. coli hosts.

XX Sequence 124 AA;

Query Match 1.2%; Score 7; DB 18; Length 124;

Best Local Similarity 100.0%; Pred. No. 93;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 274 VFGEPT 280

Db 39 vfgept 45

RESULT 40

AAW24654

ID AAW24654 standard; protein; 124 AA.

XX AAW24654;

XX 11-AUG-1997 (first entry)

XX H. pylori cytoplasmic protein 3261306.aa.

XX Transmembrane; cytoplasmic; cell envelope; flagella; transport;
 KW secreted; periplasmic; chronic gastritis; duodenal ulcer disease;
 KW activator; inhibitor; bacterial life cycle; vaccine; immunise;
 KW detection; antisense; inhibition.

XX Helicobacter pylori.

XX Key Location/Qualifiers

FT Key

FT Misc-difference 72

FT /note= "encoded by RGC"

FT Misc-difference 73

FT /note= "encoded by GTW"

XX

PN WO9719098-A1.
 XX 29-MAY-1997.
 XX 15-NOV-1996; 96WO-US18542.
 XX 17-NOV-1995; 95US-0561469.
 PR (ASTR) ASTRA AB.
 XX Smith DH;
 XX WPI; 1997-298052/27.
 DR N-PSDB; AAT77472.
 XX Helicobacter pylori nucleic acid sequences and related proteins -
 used for diagnostics and therapeutics
 XX Claim 18; Page 173; 235pp; English.
 XX This sequence represents an H. pylori cytoplasmic protein involved in
 cofactor metabolism.
 CC Helicobacter pylori has been strongly linked to chronic gastritis and
 duodenal ulcer disease. The nucleic acid sequences of the invention
 are used to evaluate compounds, especially activators or inhibitors of
 bacterial life cycle, for the ability to bind an H. pylori nucleic acid
 sequence. The nucleic acid sequences, and corresponding proteins, are
 also useful for generating vaccines for immunising subjects against H.
 pylori or for use in detecting the presence of Helicobacter species in
 a sample. Antisense nucleic acid sequences of these sequences are
 used to inhibit expression of a gene from Helicobacter species. H.
 pylori whole genomic DNA was isolated and nebulised to a median size of
 2000 bp. Purified DNA fragments were blunt-ended and ligated to unique
 BstXI-linker adapters in 100-1000 fold molar excess. These linkers are
 complementary to the BstXI-cut pMPX vectors, while the overhang is not
 self-complementary. Therefore the linkers will not concatamerise nor
 will the cut vector re-ligate itself easily. The linker-adaptor inserts
 were ligated to each of the 20 pMPX vectors to construct a series of
 shotgun subclone libraries. The purified DNA samples were then
 sequenced.
 CC Note: The ORF/protein reference number for this sequence was obtained
 from the related specification, WO9640893.
 XX Sequence 124 AA;
 SQ
 Query Match 1.2%; Score 7; DB 18; Length 124;
 Best Local Similarity 100.0%; Pred. No. 93;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 274 VFGPEPIT 280
 Db 39 vfgepit 45
 |||||
 RESULT 41
 ID AAR62918 standard; protein; 125 AA.
 XX AC AAR62918;
 XX 18-JUL-1995 (first entry)
 XX Human cytomegalovirus antibody heavy chain variable region.
 XX Human cytomegalovirus; antibody heavy chain variable region;
 KW HMCV; immunoassay; immunotherapy.
 XX Homo sapiens.
 XX WO9425490-A.
 PD 10-NOV-1994.
 PF 29-APR-1994; 94WO-US04705.
 PR 30-APR-1993; 93US-0055985.
 XX (SCRI) SCRIPPS RES INST.
 XX Barbas C, Burion R, Burton DR, Williamson A, Burioni R;
 WPI; 1994-358194/44.
 XX Human monoclonal antibodies (MABs) against human cytomegalovirus
 - also nucleic acids and cell lines producing the MABs, useful in
 diagnosis and immunotherapy
 OS Homo sapiens.
 XX WO9425490-A.
 PD 10-NOV-1994.
 PF 29-APR-1994; 94WO-US04705.
 PR 30-APR-1993; 93US-0055985.
 XX (SCRI) SCRIPPS RES INST.
 XX Barbas C, Burion R, Burton DR, Williamson A, Burioni R;
 WPI; 1994-358194/44.
 XX Human monoclonal antibodies (MABs) against human cytomegalovirus
 - also nucleic acids and cell lines producing the MABs, useful in
 diagnosis and immunotherapy
 PS Claim 1; Page 149; 171pp; English.
 XX AAR62618-R62920 and AAR62921-R62923 are human cytomegalovirus (HMCV)
 antibody heavy chain and light chain variable regions respectively.
 CC A monoclonal antibody (Mab) containing a pair of these regions can be
 used to detect HCMV, and anti-human HCMV Abs in human patients via
 a competitive immunoassay. The Mabs may also be useful in
 immunotherapy.
 XX Sequence 125 AA;
 SQ
 Query Match 1.2%; Score 7; DB 15; Length 125;
 Best Local Similarity 100.0%; Pred. No. 94;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 537 SGFSLSG 543
 Db 21 sgfsisg 27
 |||||
 RESULT 42
 ID AAR62920 standard; protein; 125 AA.
 XX AC AAR62920;
 XX 18-JUL-1995 (first entry)
 XX Human cytomegalovirus antibody heavy chain variable region.
 XX Human cytomegalovirus; antibody heavy chain variable region;
 KW HMCV; immunoassay; immunotherapy.
 XX Homo sapiens.
 XX WO9425490-A.
 PD 10-NOV-1994.
 PF 29-APR-1994; 94WO-US04705.
 PR 30-APR-1993; 93US-0055985.
 XX (SCRI) SCRIPPS RES INST.
 XX Barbas C, Burion R, Burton DR, Williamson A, Burioni R;
 WPI; 1994-358194/44.
 XX Human monoclonal antibodies (MABs) against human cytomegalovirus
 - also nucleic acids and cell lines producing the MABs, useful in
 diagnosis and immunotherapy
 PS Claim 1; Page 149; 171pp; English.
 XX AAR62618-R62920 and AAR62921-R62923 are human cytomegalovirus (HMCV)
 antibody heavy chain and light chain variable regions respectively.

XX 29-APR-1994; 94WO-US04705.
 XX 30-APR-1993; 93US-0055985.
 XX (SCRI) SCRIPPS RES INST.
 XX Barbas C, Burion R, Burton DR, Williamson A, Burioni R;
 WPI; 1994-358194/44.
 XX Human monoclonal antibodies (MABs) against human cytomegalovirus
 - also nucleic acids and cell lines producing the MABs, useful in
 diagnosis and immunotherapy
 PS Claim 1; Page 148; 171pp; English.
 XX AAR62618-R62920 and AAR62921-R62923 are human cytomegalovirus (HMCV)
 antibody heavy chain and light chain variable regions respectively.
 CC A monoclonal antibody (Mab) containing a pair of these regions can be
 used to detect HCMV, and anti-human HCMV Abs in human patients via
 a competitive immunoassay. The Mabs may also be useful in
 immunotherapy.
 XX Sequence 125 AA;
 SQ
 Query Match 1.2%; Score 7; DB 15; Length 125;
 Best Local Similarity 100.0%; Pred. No. 94;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 537 SGFSLSG 543
 Db 21 sgfsisg 27
 |||||
 RESULT 42
 ID AAR62920 standard; protein; 125 AA.
 XX AC AAR62920;
 XX 18-JUL-1995 (first entry)
 XX Human cytomegalovirus antibody heavy chain variable region.
 XX Human cytomegalovirus; antibody heavy chain variable region;
 KW HMCV; immunoassay; immunotherapy.
 XX Homo sapiens.
 XX WO9425490-A.
 PD 10-NOV-1994.
 PF 29-APR-1994; 94WO-US04705.
 PR 30-APR-1993; 93US-0055985.
 XX (SCRI) SCRIPPS RES INST.
 XX Barbas C, Burion R, Burton DR, Williamson A, Burioni R;
 WPI; 1994-358194/44.
 XX Human monoclonal antibodies (MABs) against human cytomegalovirus
 - also nucleic acids and cell lines producing the MABs, useful in
 diagnosis and immunotherapy
 PS Claim 1; Page 149; 171pp; English.
 XX AAR62618-R62920 and AAR62921-R62923 are human cytomegalovirus (HMCV)
 antibody heavy chain and light chain variable regions respectively.

CC A monoclonal antibody (Mab) containing a pair of these regions can be
 CC used to detect HCMV, and anti-human HCMV Abs in human patients via
 CC a competitive immunoassay. The Mabs may also be useful in
 CC immunotherapy.

XX SQ Sequence 125 AA;

Query Match 1.2%; Score 7; DB 15; Length 125;

Best Local Similarity 100.0%; Pred. No. 94;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 537 SGFSLSG 543

Db 21 sgfslsq 27

RESULT 43

AAR36700
 ID AAR36700 standard; Protein; 126 AA.

XX AC AAR36700;

XX DT 20-AUG-1993 (first entry)

XX DE Flagellin (amino acids 137-262).

XX KW CMP-KDO synthetase; pB776; pB410; capture; immunodot;

XX KW assay; Treponema pallidum; syphilis;

XX KW CTP: CMP-3-deoxy-manno-octulosonate cytidyl transferase.

XX OS Borrelia burgdorferi.

XX PN WO9308208-A.

XX PD 29-APR-1993.

XX PF 21-OCT-1992; 92WO-US09199.

XX PR 21-OCT-1991; 91US-0779704.

XX PA (ABBO) ABBOTT LAB.

XX PI Hunt JC, Pilot-Matias TJ, Robinson JM;

XX DR WPI; 1993-152418/18.

XX DR N-PSDB; AAQ41495.

XX PT Synthetic polypeptide immuno-reactive with Borrelia burgdorferi

XX PT antibodies - for diagnosis of Lyme disease

XX PS Claim 1; Page 48-49; 76pp; English.

XX CC Expression vector pB410 encodes fusion protein CKS-410, consisting

XX CC of 239 amino acids of CKS, 11 amino acids contributed by linker DNA

XX CC sequences, and amino acids 137 to 262 of B. burgdorferi flagellin.

XX CC The recombinant fusion proteins p776 and p410 are differentiating

XX CC polypeptides which can be used in capture assays and immunodot

XX CC assays for the detection of B. burgdorferi infection such as in

XX CC Lyme disease, without false positives due to Treponema pallidum

XX CC infection such as in patients with syphilis.

XX SQ Sequence 126 AA;

Query Match 1.2%; Score 7; DB 14;

Best Local Similarity 100.0%; Pred. No. 94;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 540 SLSGSQA 546

Db 27 slsgsqga 33

RESULT 44
 AAG24038

ID AAG24038 standard; Protein; 134 AA.

XX AC AAG24038;

XX DT 17-OCT-2000 (first entry)

XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 27562.

XX KW protein identification; signal transduction pathway; metabolic pathway;

XX KW hybridisation assay; genetic mapping; gene expression control; promoter;

XX KW termination sequence.

XX OS Arabidopsis thaliana.

XX PN EP1033405-A2.

XX PD 06-SEP-2000.

XX PF 25-FEB-2000; 2000EP-0301439.

XX PR 25-FEB-1999; 99US-0121825.

XX PR 05-MAR-1999; 99US-0123180.

XX PR 09-MAR-1999; 99US-0123548.

XX PR 23-MAR-1999; 99US-0125788.

XX PR 25-MAR-1999; 99US-0126264.

XX PR 29-MAR-1999; 99US-0126785.

XX PR 01-APR-1999; 99US-0127462.

XX PR 06-APR-1999; 99US-0128234.

XX PR 08-APR-1999; 99US-0128714.

XX PR 16-APR-1999; 99US-0129845.

XX PR 19-APR-1999; 99US-0130077.

XX PR 21-APR-1999; 99US-0130449.

XX PR 23-APR-1999; 99US-0130510.

XX PR 28-APR-1999; 99US-0130891.

XX PR 30-APR-1999; 99US-0131449.

XX PR 30-APR-1999; 99US-0132048.

XX PR 04-MAY-1999; 99US-0132484.

XX PR 05-MAY-1999; 99US-0132485.

XX PR 06-MAY-1999; 99US-0132486.

XX PR 06-MAY-1999; 99US-0132487.

XX PR 07-MAY-1999; 99US-0132863.

XX PR 11-MAY-1999; 99US-0134256.

XX PR 14-MAY-1999; 99US-0134218.

XX PR 14-MAY-1999; 99US-0134219.

XX PR 14-MAY-1999; 99US-0134221.

XX PR 14-MAY-1999; 99US-0134370.

XX PR 18-MAY-1999; 99US-0134768.

XX PR 19-MAY-1999; 99US-0134941.

XX PR 20-MAY-1999; 99US-0135124.

XX PR 21-MAY-1999; 99US-0135353.

XX PR 24-MAY-1999; 99US-0135629.

XX PR 25-MAY-1999; 99US-0136021.

XX PR 27-MAY-1999; 99US-0136392.

XX PR 28-MAY-1999; 99US-0136782.

XX PR 01-JUN-1999; 99US-0137222.

XX PR 03-JUN-1999; 99US-0137528.

XX PR 04-JUN-1999; 99US-0137502.

XX PR 07-JUN-1999; 99US-0137724.

XX PR 08-JUN-1999; 99US-0138094.

XX PR 10-JUN-1999; 99US-0138540.

XX PR 10-JUN-1999; 99US-0138847.

XX PR 14-JUN-1999; 99US-0139119.

XX PR 16-JUN-1999; 99US-0139452.

XX PR 16-JUN-1999; 99US-0139453.

XX PR 17-JUN-1999; 99US-0139492.

XX PR 18-JUN-1999; 99US-0139454.

XX PR 18-JUN-1999; 99US-0139455.

XX PR 18-JUN-1999; 99US-0139456.

XX PR 18-JUN-1999; 99US-0139457.

PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0144086.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 23-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145921.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.

PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151085.
PR 27-AUG-1999; 99US-0151086.
PR 27-AUG-1999; 99US-0151088.
PR 30-AUG-1999; 99US-0151080.
PR 31-AUG-1999; 99US-0151303.
PR 01-SEP-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 28-SEP-1999; 99US-0155659.
PR 29-SEP-1999; 99US-0156458.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 1.2%; Score 7; DB 21; Length 134;
Best Local Similarity 100.0%; Pred. No. 99;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 152 LEALVKE 158
Db 64 lealvke 70
|||||

RESULT 45

AA43865

ID AA43865 standard; Peptide; 139 AA.

XX

AC AA43865;

XX

DT 11-FEB-2000 (first entry)
 XX Heavy chain (VH) gene sequence of Bacillus anthracis antibody 9-1.
 DE Heavy chain; VH; IgG; monoclonal antibody; spore; Bacillus;
 XX VH gene usage; anthrax.
 KW Mus sp.
 XX WO9955842-A1.
 PN 04-NOV-1999.
 PD 27-APR-1999; 99WO-US09122.
 XX 29-APR-1998; 98US-0069628.
 XX (UABR-) UAB RES FOUND.
 XX Kearney JF;
 PI WPI; 2000-013435/01.
 XX Monoclonal antibody specific for Bacillus spores, used to detect
 PT anthrax -
 XX Example 13; Page 52-53; 64pp; English.
 PS AAY43862-75 represent the amino acid sequences of the heavy chain
 XX (VH) gene sequences from antibodies against Bacillus anthracis
 CC spores. The antibodies are produced by different hybridomas. The
 CC specification describes monoclonal antibodies (especially IgG
 CC antibodies) which are highly specific and can discriminate between the
 CC spores of the Bacillus family. The antibodies are produced by exposing
 CC mice to Bacillus spores. The humoral immune response to Bacillus spores
 CC shows a conservation of VH gene usage which is distinct for each spore.
 CC Peptide fragments derived from the antibodies are also capable of
 CC binding spores. The monoclonal antibody, and peptide fragments of it, can
 CC be used to detect Bacillus spores in a field sample. It is particularly
 CC uses for detecting anthrax in a field sample.
 XX SQ Sequence 139 AA;

Query Match 1.28; Score 7; DB 21; Length 139;
 Best Local Similarity 100.0%; Pred. No. 1e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 537 SGFSLSG 543
 DB 12 sgfsisg 18
 |||||

Search completed: July 30, 2001, 11:42:55
 Job time: 139 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2001, 11:41:01 ; Search time 12.41 Seconds
(without alignments)
949.047 Million cell updates/sec

Title: US-09-367-496-8
Perfect score: 572
Sequence: 1 MSFGKKSIPRITSRLIR.....RTAQKIMAPGGRSNTLSLS 572

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 197339 seqs, 20590346 residues

Word size : 0

Total number of hits satisfying chosen parameters: 197339

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/1/iaa/PTUS_COMB.pep.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	10	1.7	460	1 US-08-289-709-1	Sequence 1, Appl
2	10	1.7	460	1 US-08-602-656-1	Sequence 1, Appl
3	7	1.2	126	1 US-08-500-222-8	Sequence 8, Appl
4	7	1.2	126	1 US-08-500-125-8	Sequence 8, Appl
5	7	1.2	126	1 US-07-779-704B-8	Sequence 8, Appl
6	7	1.2	248	1 US-08-500-222-4	Sequence 4, Appl
7	7	1.2	248	1 US-08-500-125-4	Sequence 4, Appl
8	7	1.2	248	1 US-07-779-704B-4	Sequence 4, Appl
9	7	1.2	336	1 US-07-641-143B-2	Sequence 2, Appl
10	7	1.2	336	1 US-08-124-290-2	Sequence 2, Appl
11	7	1.2	336	3 US-08-696-372A-2	Sequence 2, Appl
12	7	1.2	336	4 US-08-235-836C-52	Sequence 52, Appl
13	7	1.2	336	4 US-08-235-836C-54	Sequence 54, Appl
14	7	1.2	341	4 US-09-196-293-14	Sequence 14, Appl
15	7	1.2	341	4 US-08-209-603E-14	Sequence 14, Appl
16	7	1.2	376	1 US-08-500-222-6	Sequence 6, Appl
17	7	1.2	376	1 US-08-500-125-6	Sequence 6, Appl
18	7	1.2	376	2 US-07-779-704B-6	Sequence 6, Appl
19	7	1.2	464	1 US-07-688-352C-16	Sequence 16, Appl
20	7	1.2	464	2 US-08-474-379C-16	Sequence 16, Appl
21	7	1.2	464	3 US-09-146-249A-16	Sequence 16, Appl
22	7	1.2	464	3 US-08-206-188B-16	Sequence 16, Appl
23	7	1.2	464	5 PCT-US91-02714-16	Sequence 16, Appl
24	7	1.2	498	1 US-08-500-222-2	Sequence 2, Appl
25	7	1.2	498	1 US-08-500-125-2	Sequence 2, Appl
26	7	1.2	498	2 US-07-779-704B-2	Sequence 2, Appl
27	7	1.2	514	4 US-08-796-899-25	Sequence 25, Appl

28 6 1.0 8 3 US-09-188-579-62 Sequence 62, Appl
29 6 1.0 8 4 US-08-960-780-69 Sequence 69, Appl
30 6 1.0 8 4 US-09-315-444-62 Sequence 62, Appl
31 6 1.0 8 4 US-09-073-898-69 Sequence 69, Appl
32 6 1.0 10 3 US-08-159-339A-1037 Sequence 1037, Ap
33 6 1.0 15 2 US-08-480-190-162 Sequence 162, App
34 6 1.0 15 2 US-08-488-379-162 Sequence 162, App
35 6 1.0 15 5 PCT-US93-07545-162 Sequence 162, App
36 6 1.0 17 2 US-08-480-190-161 Sequence 161, App
37 6 1.0 17 2 US-08-488-379-161 Sequence 161, App
38 6 1.0 17 5 PCT-US93-07545-161 Sequence 161, App
39 6 1.0 17 5 PCT-US95-08596-21 Sequence 21, Appl
40 6 1.0 17 5 PCT-US95-08621-20 Sequence 20, Appl
41 6 1.0 18 2 US-08-480-190-160 Sequence 160, App
42 6 1.0 18 2 US-08-488-379-160 Sequence 160, App
43 6 1.0 18 5 PCT-US93-07545-160 Sequence 160, App
44 6 1.0 20 1 US-08-468-543-10 Sequence 10, Appl
45 6 1.0 20 2 US-08-469-692-10 Sequence 10, Appl

ALIGNMENTS

RESULT 1
US-08-289-709-1
; Sequence 1, Application US/08289709
; Patent No. 5523224
; GENERAL INFORMATION:
; APPLICANT: Burtcher, Helmut; Lang, Gunter; Popp, Friedrich
; TITLE OF INVENTION: Recombinant D-Hydantoinase, A Process for the
; TITLE OF INVENTION: Production and Use
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/289,709
; FILING DATE: 12-AUGUST-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: P 43 28 829.4
; FILING DATE: 27-AUGUST-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, No. 5523224man D.
; REGISTRATION NUMBER: 30,946
; REFERENCE/DOCKET NUMBER: BOER 1041
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 460 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-289-709-1

Query Match 1.7%; Score 10; DB 1; Length 460;
Best Local Similarity 100.0%; Pred. No. 0.063;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 402 AVGSDADIVI 411
|||||

Db 386 AVGSDADLVI 395

RESULT 2
US-08-602-656-1
; Sequence 1, Application US/08602656
; Patent No. 5679571
; GENERAL INFORMATION:
; APPLICANT: Burtischer, Helmut; Lang, Gunter; Popp, Friedrich
; TITLE OF INVENTION: Recombinant D-Hydantoinase, A Process For the
; TITLE OF INVENTION: Production and Use
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/602,656
; FILING DATE: 16-FEB-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/289,709
; FILING DATE: 12-AUGUST-1994
; APPLICATION NUMBER: P 43 28 829.4
; FILING DATE: 27-AUGUST-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, No. 5679571man D.
; REGISTRATION NUMBER: 30,946
; REFERENCE/DOCKET NUMBER: BOER 1041
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 460 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-602-656-1

Query Match 1.7%; Score 10; DB 1; Length 460;
Best Local Similarity 100.0%; Pred. No. 0.063;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 402 AVGSDADLVI 411
|||||||
Db 386 AVGSDADLVI 395

RESULT 3
US-08-500-222-8
; Sequence 8, Application US/08500222
; Patent No. 5643733
; GENERAL INFORMATION:
; APPLICANT: Robinson, John M
; APPLICANT: Pilot-Matias, Tami J
; APPLICANT: Hunt, Jeffrey C
; TITLE OF INVENTION: Borrelia burgdorferi antigens
; TITLE OF INVENTION: and uses thereof
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Abbott Laboratories
; STREET: One Abbott Park Road
; CITY: Abbott Park
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60064
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/500,125
; FILING DATE:

STATE: Illinois
COUNTRY: USA
ZIP: 60064
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/500,222
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/779,704
FILING DATE: 21-OCT-1991
ATTORNEY/AGENT INFORMATION:
NAME: Wong, Wean Khing
REGISTRATION NUMBER: 33561
REFERENCE/DOCKET NUMBER: 5051.US.01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 708-937-9396
TELEFAX: 708-937-9556
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 126 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: Protein
HYPOTHETICAL:
ANTI-SENSE: NO
FRAGMENT TYPE: Interval
ORIGINAL SOURCE:
ORGANISM: Borrelia burgdorferi
STRAIN: B31
US-08-500-222-8

Query Match 1.2%; Score 7; DB 1; Length 126;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 7; Conservative 0; Mismatches 0;

QY 540 SILSGSQA 546
|||||||
Db 27 SILSGSQA 33

RESULT 4
US-08-500-125-8
; Sequence 8, Application US/08500125
; Patent No. 5643751
; GENERAL INFORMATION:
; APPLICANT: Robinson, John M
; APPLICANT: Pilot-Matias, Tami J
; APPLICANT: Hunt, Jeffrey C
; TITLE OF INVENTION: Borrelia burgdorferi antigens
; TITLE OF INVENTION: and uses thereof
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Abbott Laboratories
; STREET: One Abbott Park Road
; CITY: Abbott Park
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60064
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/500,125
; FILING DATE:

;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/779,704
;; FILING DATE: 21-OCT-1991
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Wong, Wean Khing
;; REGISTRATION NUMBER: 33561
;; REFERENCE/DOCKET NUMBER: 5051.US.01
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 708-937-9396
;; TELEFAX: 708-937-9556
;; INFORMATION FOR SEQ ID NO: 8:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 126 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: Protein
;; HYPOTHETICAL:
;; ANTI-SENSE: NO
;; FRAGMENT TYPE: interval
;; ORIGINAL SOURCE:
;; ORGANISM: Borrelia burgdorferi
;; STRAIN: B31
;; US-08-500-125-8

Query Match 1.2%; Score 7; DB 1; Length 126;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 540 SLGSQA 546
Db 27 SLGSQA 33

RESULT 5
US-07-779-704B-8
; Sequence 8, Application US/07779704B
; Patent No. 5965702
; GENERAL INFORMATION:
; APPLICANT: Robinson, John M
; APPLICANT: Pilot-Matias, Tami J
; APPLICANT: Hunt, Jeffrey C
; TITLE OF INVENTION: Borrelia burgdorferi antigens
; TITLE OF INVENTION: and uses thereof
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Abbott Laboratories
; STREET: One Abbott Park Road
; CITY: Abbott Park
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60064
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/779,704B
; FILING DATE: 19911021
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Wong, Wean Khing
; REGISTRATION NUMBER: 33561
; REFERENCE/DOCKET NUMBER: 5051.US.01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 708-937-9396
; TELEFAX: 708-937-9556
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 126 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear

;; MOLECULE TYPE: Protein
;; HYPOTHETICAL:
;; ANTI-SENSE: NO
;; FRAGMENT TYPE: interval
;; ORIGINAL SOURCE:
;; ORGANISM: Borrelia burgdorferi
;; STRAIN: B31
;; US-07-779-704B-8

Query Match 1.2%; Score 7; DB 2; Length 126;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 540 SLGSQA 546
Db 27 SLGSQA 33

RESULT 6
US-08-500-222-4
; Sequence 4, Application US/08500222
; Patent No. 5643733
; GENERAL INFORMATION:
; APPLICANT: Robinson, John M
; APPLICANT: Pilot-Matias, Tami, J
; APPLICANT: Hunt, Jeffrey C
; TITLE OF INVENTION: Borrelia burgdorferi antigens
; TITLE OF INVENTION: and uses thereof
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Abbott Laboratories
; STREET: One Abbott Park Road
; CITY: Abbott Park
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60064
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/500,222
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/779,704
; FILING DATE: 21-OCT-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Wong, Wean Khing
; REGISTRATION NUMBER: 33561
; REFERENCE/DOCKET NUMBER: 5051.US.01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 708-937-9396
; TELEFAX: 708-937-9556
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 248 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: internal
; ORIGINAL SOURCE:
; ORGANISM: Borrelia burgdorferi
; STRAIN: B31
; US-08-500-222-4

Query Match 1.2%; Score 7; DB 1; Length 248;
Best Local Similarity 100.0%; Pred. No. 44;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 540 SLSGSQA 546
 Db 100 SLSGSQA 106

RESULT 7

US-08-500-125-4
 ; Sequence 4, Application US/08500125
 ; Patent No. 5643751

; GENERAL INFORMATION:
 ; APPLICANT: Robinson, John M
 ; APPLICANT: Pilot-Matias, Tami J
 ; APPLICANT: Hunt, Jeffrey C
 ; TITLE OF INVENTION: Borrelia burgdorferi antigens
 ; TITLE OF INVENTION: and uses thereof
 ; NUMBER OF SEQUENCES: 16
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Abbott Laboratories
 ; STREET: One Abbott Park Road
 ; CITY: Abbott Park
 ; STATE: Illinois
 ; COUNTRY: USA
 ; ZIP: 60064

; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/500.125
 ; FILING DATE:
 ; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/779,704
 ; FILING DATE: 21-OCT-1991

; ATTORNEY/AGENT INFORMATION:
 ; NAME: Wong, Wean Khing
 ; REGISTRATION NUMBER: 33561
 ; REFERENCE/DOCKET NUMBER: 5051.US.01
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 708-937-9396
 ; TELEFAX: 708-937-9556

; INFORMATION FOR SEQ ID NO: 4:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 248 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear

; MOLECULE TYPE: protein
 ; HYPOTHETICAL: NO
 ; ANTI-SENSE: NO
 ; FRAGMENT TYPE: internal
 ; ORIGINAL SOURCE:
 ; ORGANISM: Borrelia burgdorferi
 ; STRAIN: B31

US-08-500-125-4

Query Match 1.2%; Score 7; DB 1; Length 248;
 Best Local Similarity 100.0%; Pred. No. 44;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 540 SLSGSQA 546
 Db 100 SLSGSQA 106

RESULT 8

US-07-779-704B-4
 ; Sequence 4, Application US/07779704B
 ; Patent No. 5965702

; GENERAL INFORMATION:

; APPLICANT: Robinson, John M
 ; APPLICANT: Pilot-Matias, Tami J
 ; APPLICANT: Hunt, Jeffrey C
 ; TITLE OF INVENTION: Borrelia burgdorferi antigens
 ; TITLE OF INVENTION: and uses thereof
 ; NUMBER OF SEQUENCES: 16
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Abbott Laboratories
 ; STREET: One Abbott Park Road
 ; CITY: Abbott Park
 ; STATE: Illinois
 ; COUNTRY: USA
 ; ZIP: 60064

; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/07/779,704B
 ; FILING DATE: 19911021
 ; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:
 ; NAME: Wong, Wean Khing
 ; REGISTRATION NUMBER: 33561
 ; REFERENCE/DOCKET NUMBER: 5051.US.01
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 708-937-9396
 ; TELEFAX: 708-937-9556

; INFORMATION FOR SEQ ID NO: 4:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 248 amino acids
 ; TYPE: AMINO ACID
 ; TOPOLOGY: linear

; MOLECULE TYPE: protein
 ; HYPOTHETICAL: NO
 ; ANTI-SENSE: NO
 ; FRAGMENT TYPE: internal
 ; ORIGINAL SOURCE:
 ; ORGANISM: Borrelia burgdorferi
 ; STRAIN: B31

US-07-779-704B-4

Query Match 1.2%; Score 7; DB 2; Length 248;
 Best Local Similarity 100.0%; Pred. No. 44;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 540 SLSGSQA 546
 Db 100 SLSGSQA 106

RESULT 9

US-07-641-143B-2
 ; Sequence 2, Application US/07641143B
 ; Patent No. 5436000

; GENERAL INFORMATION:
 ; APPLICANT: Barbour, Alan G.
 ; APPLICANT: Bundoc, Virgilio
 ; TITLE OF INVENTION: Flagella-less Borrelia
 ; NUMBER OF SEQUENCES: 3
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Arnold, White & Durkee
 ; STREET: P.O. Box 4433
 ; CITY: Houston
 ; STATE: TX
 ; COUNTRY: USA
 ; ZIP: 77210

; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS

;; SOFTWARE: PatentIn Release #1.0, Version #1.25

;; CURRENT APPLICATION DATA:

;; APPLICATION NUMBER: US/07/641,143B

;; FILING DATE: 11-JAN-1991

;; CLASSIFICATION: 424

;; ATTORNEY/AGENT INFORMATION:

;; NAME: Patterson, Melinda

;; REGISTRATION NUMBER: 33,062

;; REFERENCE/DOCKET NUMBER: UTSK092

;; TELECOMMUNICATION INFORMATION:

;; TELEPHONE: 713-787-1592

;; TELEFAX: 713-789-2679

;; INFORMATION FOR SEQ ID NO: 2:

;; SEQUENCE CHARACTERISTICS:

;; LENGTH: 336 amino acids

;; TYPE: amino acid

;; TOPOLOGY: linear

;; MOLECULE TYPE: protein

US-07-641-143B-2

Query Match 1.2%; Score 7; DB 1; Length 336;

Best Local Similarity 100.0%; Pred. No. 59;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 540 SLSGSQA 546

Db 163 SLSGSQA 169

RESULT 10

US-08-124-290-2

;; Sequence 2, Application US/08124290

;; Patent No. 5585102

;; GENERAL INFORMATION:

;; APPLICANT: Barbour, Alan G.

;; APPLICANT: Bundoc, Virgilio

;; TITLE OF INVENTION: Flagella-less Borrelia

;; NUMBER OF SEQUENCES: 3

;; CORRESPONDENCE ADDRESS:

;; ADDRESSEE: Arnold, White & Durkee

;; STREET: P.O. Box 4433

;; CITY: Houston

;; STATE: TX

;; COUNTRY: USA

;; ZIP: 77210

;; COMPUTER READABLE FORM:

;; MEDIUM TYPE: Floppy disk

;; COMPUTER: IBM PC compatible

;; OPERATING SYSTEM: PC-DOS/MS-DOS

;; SOFTWARE: PatentIn Release #1.0, Version #1.25

;; CURRENT APPLICATION DATA:

;; APPLICATION NUMBER: US/08/124,290

;; FILING DATE:

;; CLASSIFICATION: 424

;; PRIOR APPLICATION DATA:

;; APPLICATION NUMBER: US/07/641,143

;; FILING DATE: 11-JAN-1991

;; CLASSIFICATION: 424

;; ATTORNEY/AGENT INFORMATION:

;; NAME: Patterson, Melinda

;; REGISTRATION NUMBER: 33,062

;; REFERENCE/DOCKET NUMBER: UTSK092

;; TELECOMMUNICATION INFORMATION:

;; TELEPHONE: 713-787-1592

;; TELEFAX: 713-789-2679

;; INFORMATION FOR SEQ ID NO: 2:

;; SEQUENCE CHARACTERISTICS:

;; LENGTH: 336 amino acids

;; TYPE: amino acid

;; TOPOLOGY: linear

;; MOLECULE TYPE: protein

US-08-124-290-2

Query Match 1.2%; Score 7; DB 1; Length 336;

Best Local Similarity 100.0%; Pred. No. 59;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 540 SLSGSQA 546

Db 163 SLSGSQA 169

RESULT 11

US-08-696-372A-2

;; Sequence 2, Application US/08696372A

;; Patent No. 6077515

;; GENERAL INFORMATION:

;; APPLICANT: Barbour, Alan G.

;; APPLICANT: Bundoc, Virgilio

;; TITLE OF INVENTION: Flagella-less Borrelia

;; NUMBER OF SEQUENCES: 3

;; CORRESPONDENCE ADDRESS:

;; ADDRESSEE: Arnold, White & Durkee

;; STREET: PO Box 4433

;; CITY: Houston

;; STATE: Texas

;; COUNTRY: USA

;; ZIP: 77210

;; COMPUTER READABLE FORM:

;; MEDIUM TYPE: Floppy disk

;; COMPUTER: IBM PC compatible

;; OPERATING SYSTEM: PC-DOS/MS-DOS

;; SOFTWARE: PatentIn Release #1.0, Version #1.25

;; CURRENT APPLICATION DATA:

;; APPLICATION NUMBER: US/08/696,372A

;; FILING DATE:

;; CLASSIFICATION:

;; PRIOR APPLICATION DATA:

;; APPLICATION NUMBER: US/07/641,143

;; FILING DATE: 11-JAN-1991

;; ATTORNEY/AGENT INFORMATION:

;; NAME: Patterson, Melinda

;; REGISTRATION NUMBER: 33,062

;; TELECOMMUNICATION INFORMATION:

;; TELEPHONE: 713-787-1592

;; TELEFAX: 713-789-2679

;; INFORMATION FOR SEQ ID NO: 2:

;; SEQUENCE CHARACTERISTICS:

;; LENGTH: 336 amino acids

;; TYPE: amino acid

;; TOPOLOGY: linear

;; MOLECULE TYPE: protein

US-08-696-372A-2

Query Match

1.2%; Score 7; DB 3; Length 336;

Best Local Similarity 100.0%; Pred. No. 59;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 540 SLSGSQA 546

Db 163 SLSGSQA 169

RESULT 12

US-08-235-836C-52

;; Sequence 52, Application US/08235836C

;; Patent No. 6248562

;; GENERAL INFORMATION:

;; APPLICANT: Dunn, John J.

;; APPLICANT: Luft, Benjamin J.

;; TITLE OF INVENTION: NO. 6248562e1 Chimeric Proteins Comprising

;; TITLE OF INVENTION: Borrelia Polypeptides and Uses Therefor

;; NUMBER OF SEQUENCES: 144

;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Brookhaven National Laboratory
;; STREET:
;; CITY: Upton
;; STATE: NY
;; COUNTRY: USA
;; ZIP: 11973
;;
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/235,836C
;; FILING DATE: 29-APR-1994
;; CLASSIFICATION: 435
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/148,191
;; FILING DATE: 01-11-93
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Bogosian, Margaret C.
;; REGISTRATION NUMBER: 25,324
;; REFERENCE/DOCKET NUMBER: BNL93-28A
;; TELEPHONE: (516) 282-7338
;; TELEFAX: (516) 282-3729
;; INFORMATION FOR SEQ ID NO: 52:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 336 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;; US-08-235-836C-52

Query Match 1.2%; Score 7; DB 4; Length 336;
Best Local Similarity 100.0%; Pred. No. 59;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 540 SLGSGQA 546
DB 163 SLGSGQA 169

RESULT 13
US-08-235-836C-54
; Sequence 54, Application US/08235836C
; Patent No. 6248562
; GENERAL INFORMATION:
; APPLICANT: Dunn, John J.
; TITLE OF INVENTION: Chimeric Proteins Comprising
; TITLE OF INVENTION: Borrelia Polypeptides and Uses Therefor
; NUMBER OF SEQUENCES: 144
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brookhaven National Laboratory
; STREET:
; CITY: Upton
; STATE: NY
; COUNTRY: USA
; ZIP: 11973
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/235,836C
; FILING DATE: 29-APR-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/148,191
; FILING DATE: 01-11-93

;; ATTORNEY/AGENT INFORMATION:
;; NAME: Bogosian, Margaret C.
;; REGISTRATION NUMBER: 25,324
;; REFERENCE/DOCKET NUMBER: BNL93-28A
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (516) 282-7338
;; TELEFAX: (516) 282-3729
;; INFORMATION FOR SEQ ID NO: 54:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 336 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;; US-08-235-836C-54

Query Match 1.2%; Score 7; DB 4; Length 336;
Best Local Similarity 100.0%; Pred. No. 59;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 540 SLGSGQA 546
DB 163 SLGSGQA 169

RESULT 14
US-09-196-293-14
; Sequence 14, Application US/09196293
; Patent No. 6183755
; GENERAL INFORMATION:
; APPLICANT: Fuchs, Renate
; APPLICANT: Motz, Manfred
; APPLICANT: Soutscheck, Erwin
; APPLICANT: Wilske, Bettina
; APPLICANT: Preac-Mursic, Vera
; TITLE OF INVENTION: Active proteins from Borrelia
; TITLE OF INVENTION: burgdorferi
; FILE REFERENCE: 738.001US2
; CURRENT APPLICATION NUMBER: US/09/196,293
; CURRENT FILING DATE: 1998-11-19
; EARLIER APPLICATION NUMBER: US 08/209,603
; EARLIER FILING DATE: 1994-03-10
; EARLIER APPLICATION NUMBER: US 07/862,535
; EARLIER FILING DATE: 1992-06-19
; EARLIER APPLICATION NUMBER: WO PCT/EP90/02282
; EARLIER FILING DATE: 1990-12-21
; EARLIER APPLICATION NUMBER: DE P39 42 728.5
; EARLIER FILING DATE: 1989-12-22
; EARLIER APPLICATION NUMBER: DE P40 18 988.0
; EARLIER FILING DATE: 1990-06-13
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 341
; TYPE: PRT
; ORGANISM: Borrelia burgdorferi
; US-09-196-293-14

Query Match 1.2%; Score 7; DB 4; Length 341;
Best Local Similarity 100.0%; Pred. No. 60;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 540 SLGSGQA 546
DB 168 SLGSGQA 174

RESULT 15
US-08-209-603E-14
; Sequence 14, Application US/08209603E
; Patent No. 6248538
; GENERAL INFORMATION:

APPLICANT: FUCHS, RENATE
APPLICANT: WILSKE, BETTINA
APPLICANT: PREAC-MURSKI, VERA
APPLICANT: MOTZ, MANFRED
APPLICANT: SOUTSCHECK, ERWIN
TITLE OF INVENTION: IMMUNOLOGICALLY ACTIVE PROTEINS
FROM BORRELIA BURGDORFERI
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROOKS HALDT HAFNER & DELAHUNTY
STREET: 99 PARK AVENUE
CITY: NEW YORK
STATE: NY
COUNTRY: USA
ZIP: 10016
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" FLOPPY DISC
COMPUTER: AT&T - IBM COMPATIBLE
OPERATING SYSTEM: MS-DOS Version 6.2
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/209,603E
FILING DATE: 10-MAR-1994
CLASSIFICATION: 436
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/EP90/02282
FILING DATE: 21-DEC-1990
APPLICATION NUMBER: US 07/862,535
FILING DATE: 19-JUN-1992
ATTORNEY/AGENT INFORMATION:
NAME: ROBINSON, WILLIAM R.
REGISTRATION NUMBER: 27,224
REFERENCE/DOCKET NUMBER: LKR-9217-A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 697-3355
TELEFAX: (212) 557-3635
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 341
TYPE: AMINO ACID
TOPOLOGY: LINEAR
MOLECULE TYPE: PROTEIN
DESCRIPTION: N/A
HYPOTHETICAL: N/A
ANTI-SENSE: N/A
FRAGMENT TYPE: N/A
ORIGINAL SOURCE:
ORGANISM: B. BURGDORFERI
IMMEDIATE SOURCE:
LIBRARY: DSM 5662
POSITION IN GENOME: N/A
FEATURE:
IDENTIFICATION METHOD: amino acid analysis
PUBLICATION INFORMATION: N/A
US-08-209-603E-14

Query Match 1.2%; Score 7; DB 4; Length 341;
Best Local Similarity 100.0%; Pred. No. 60;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 540 SLGSQA 546
DB 168 SLGSQA 174

RESULT 16
US-08-500-222-6
Sequence 6, Application US/08500222
Patent No. 5643733
GENERAL INFORMATION:
APPLICANT: Robinson, John M
APPLICANT: Pilot-Matias, Tami J

APPLICANT: Hunt, Jeffrey C
TITLE OF INVENTION: Borrelia burgdorferi antigens
and uses thereof
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: Abbott Laboratories
STREET: One Abbott Park Road
CITY: Abbott Park
STATE: Illinois
COUNTRY: USA
ZIP: 60064
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/500,222
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/779,704
FILING DATE: 21-OCT-1991
ATTORNEY/AGENT INFORMATION:
NAME: Wong, Wean King
REGISTRATION NUMBER: 33561
REFERENCE/DOCKET NUMBER: 5051.US.01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 708-937-9396
TELEFAX: 708-937-9556
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 376 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: internal
ORIGINAL SOURCE:
ORGANISM: Borrelia burgdorferi
STRAIN: B31
US-08-500-222-6
Query Match 1.2%; Score 7; DB 1; Length 376;
Best Local Similarity 100.0%; Pred. No. 66;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 540 SLGSQA 546
DB 277 SLGSQA 283
RESULT 17
US-08-500-125-6
Sequence 6, Application US/08500125
Patent No. 5643751
GENERAL INFORMATION:
APPLICANT: Robinson, John M
APPLICANT: Pilot-Matias, Tami J
APPLICANT: Hunt, Jeffrey C
TITLE OF INVENTION: Borrelia burgdorferi antigens
and uses thereof
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: Abbott Laboratories
STREET: One Abbott Park Road
CITY: Abbott Park
STATE: Illinois
COUNTRY: USA
ZIP: 60064
COMPUTER READABLE FORM:

;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; FILING DATE:
;; CLASSIFICATION: 435
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/779,704
;; FILING DATE: 21-OCT-1991
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Wong, Wean Khing
;; REGISTRATION NUMBER: 33561
;; REFERENCE/DOCKET NUMBER: 5051.US.01
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 708-937-9396
;; TELEFAX: 708-937-9556
;; INFORMATION FOR SEQ ID NO: 6:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 376 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;; HYPOTHETICAL: NO
;; ANTI-SENSE: NO
;; FRAGMENT TYPE: internal
;; ORIGINAL SOURCE:
;; ORGANISM: Borrelia burgdorferi
;; STRAIN: B31
;; US-08-500-125-6

Query Match 1.2%; Score 7; DB 1; Length 376;
Best Local Similarity 100.0%; Pred. No. 66;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 540 SLSGSQA 546
|||||
Db 277 SLSGSQA 283

RESULT 18
US-07-779-704B-6
; Sequence 6, Application US/07779704B
; Patent No. 5965702
; GENERAL INFORMATION:
; APPLICANT: Robinson, John M
; APPLICANT: Pilot-Matias, Tam J
; APPLICANT: Hunt, Jeffrey C
; TITLE OF INVENTION: Borrelia burgdorferi antigens
; TITLE OF INVENTION: and uses thereof
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Abbott Laboratories
; STREET: One Abbott Park Road
; CITY: Abbott Park
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60064
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/779,704B
; FILING DATE: 19911021
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Wong, Wean Khing
; REGISTRATION NUMBER: 33561
; REFERENCE/DOCKET NUMBER: 5051.US.01

;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 708-937-9396
;; TELEFAX: 708-937-9556
;; INFORMATION FOR SEQ ID NO: 6:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 376 amino acids
;; TYPE: AMINO ACID
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;; HYPOTHETICAL: NO
;; ANTI-SENSE: NO
;; FRAGMENT TYPE: internal
;; ORIGINAL SOURCE:
;; ORGANISM: Borrelia burgdorferi
;; STRAIN: B31
;; US-07-779-704B-6

Query Match 1.2%; Score 7; DB 2; Length 376;
Best Local Similarity 100.0%; Pred. No. 66;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 540 SLSGSQA 546
|||||
Db 277 SLSGSQA 283

RESULT 19
US-07-688-352C-16
; Sequence 16, Application US/07688352C
; Patent No. 5527896
; GENERAL INFORMATION:
; APPLICANT: Wigler, Michael H.
; APPLICANT: Colicelli, John J.
; TITLE OF INVENTION: Cloning by Complementation and Related
; TITLE OF INVENTION: Processes
; NUMBER OF SEQUENCES: 57
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
; ADDRESSEE: Bicknell
; STREET: Two First National Plaza, 20 South Clark
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/688,352C
; FILING DATE: 19910419
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/511,715
; FILING DATE: 20-APR-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Borun, Michael F.
; REGISTRATION NUMBER: 25447
; REFERENCE/DOCKET NUMBER: 27805/30197
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 346-5750
; TELEFAX: (312) 984-9740
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 464 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-688-352C-16

```

Query Match          1.2%; Score 7; DB 1; Length 464;
Best Local Similarity 100.0%; Pred. No. 81;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 477 FVYKRIK 483
   |||||
DB 440 FVYKRIK 446

RESULT 20
US-08-474-379C-16
; Sequence 16, Application US/08474379C
; Patent No. 5977305
; GENERAL INFORMATION:
; APPLICANT: Wigler, Michael H.
; APPLICANT: Colicelli, John J.
; TITLE OF INVENTION: CLONING BY COMPLEMENTATION AND RELATED
; TITLE OF INVENTION: PROCESSES
; NUMBER OF SEQUENCES: 88
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 233 South Wacker Drive/6300 Sears Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/474,379C
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/511,715
; FILING DATE: 20-APR-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/206,188
; FILING DATE: 01-MAR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/688,352
; FILING DATE: 19-APR-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Clough, David W.
; REGISTRATION NUMBER: 36,107
; REFERENCE/DOCKET NUMBER: 27866/32771
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 474-6300
; TELEFAX: (312) 474-0448
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 464 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-474-379C-16

```

```

Query Match          1.2%; Score 7; DB 2; Length 464;
Best Local Similarity 100.0%; Pred. No. 81;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 477 FVYKRIK 483
   |||||
DB 440 FVYKRIK 446

```

```

RESULT 21
US-09-146-249A-16

```

```

; Sequence 16, Application US/09146249A
; Patent No. 6069240
; GENERAL INFORMATION:
; APPLICANT: Wigler, Michael H.
; APPLICANT: Colicelli, John J.
; TITLE OF INVENTION: Cloning by Complementation and Related
; TITLE OF INVENTION: Processes
; NUMBER OF SEQUENCES: 85
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/146,249A
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/511,715
; FILING DATE: 20-APR-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Clough, David W.
; REGISTRATION NUMBER: 36,107
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6300
; TELEFAX: 312-474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 464 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-146-249A-16

```

```

Query Match          1.2%; Score 7; DB 3; Length 464;
Best Local Similarity 100.0%; Pred. No. 81;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 477 FVYKRIK 483
   |||||
DB 440 FVYKRIK 446

```

```

RESULT 22
US-08-206-1f8B-16
; Sequence 16, Application US/08206188B
; Patent No. 6100025
; GENERAL INFORMATION:
; APPLICANT: Wigler, Michael H.
; APPLICANT: Colicelli, John J.
; TITLE OF INVENTION: Cloning by Complementation and Related
; TITLE OF INVENTION: Processes
; NUMBER OF SEQUENCES: 84
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

```

SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/206.188B
FILING DATE: 01-MAR-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/511.715
FILING DATE: 20-APR-1990
ATTORNEY/AGENT INFORMATION:
NAME: Clough, David W.
REGISTRATION NUMBER: 36107
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6300
TELEFAX: 312-474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 464 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-206-188B-16

Query Match 1.2%; Score 7; DB 3; Length 464;
Best Local Similarity 100.0%; Pred. No. 81;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 477 FVYKRIK 483
DB 440 FVYKRIK 446

RESULT 23

PCT-US91-02714-16
Sequence 16, Application PC/TUS9102714
GENERAL INFORMATION:
APPLICANT: Wigler, Michael H.
APPLICANT: Colicelli, John J.
TITLE OF INVENTION: Cloning by Complementation and Related
TITLE OF INVENTION: Processes
NUMBER OF SEQUENCES: 55
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
ADDRESSEE: Bicknell
STREET: Two First National Plaza, 20 South Clark
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60603
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US91/02714
FILING DATE: 19910419
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/511.715
FILING DATE: 20-APR-1990
ATTORNEY/AGENT INFORMATION:
NAME: Borun, Michael F.
REGISTRATION NUMBER: 25447
REFERENCE/DOCKET NUMBER: 27805/30197
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312) 346-5750
TELEFAX: (312) 984-9740
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:

LENGTH: 464 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US91-02714-16

Query Match 1.2%; Score 7; DB 5; Length 464;
Best Local Similarity 100.0%; Pred. No. 81;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 477 FVYKRIK 483
DB 440 FVYKRIK 446

RESULT 24

US-08-500-222-2
Sequence 2, Application US/08500222
Patent No. 5643733
GENERAL INFORMATION:
APPLICANT: Robinson, John M
APPLICANT: Pilot-Matias, Tami J
APPLICANT: Hunt, Jeffrey C
TITLE OF INVENTION: Borrelia burgdorferi antigens
TITLE OF INVENTION: and uses thereof
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: Abbott Laboratories
STREET: One Abbott Park Road
CITY: Abbott Park
STATE: Illinois
COUNTRY: USA
ZIP: 60064
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/500,222
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/779,704
FILING DATE: 21-OCT-1991
ATTORNEY/AGENT INFORMATION:
NAME: Wong, Wean Khing
REGISTRATION NUMBER: 33561
REFERENCE/DOCKET NUMBER: 5051.US.01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 708-937-9396
TELEFAX: 708-937-9556
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 498 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: Internal
ORIGINAL SOURCE:
ORGANISM: Borrelia burgdorferi
STRAIN: B31
US-08-500-222-2

Query Match 1.2%; Score 7; DB 1; Length 498;
Best Local Similarity 100.0%; Pred. No. 86;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 540 SLSSGSA 546

Db 350 SLGSQA 356
|||||||

RESULT 25

US-08-500-125-2
; Sequence 2, Application US/08500125
; Patent No. 5643751

GENERAL INFORMATION:

APPLICANT: Robinson, John M
APPLICANT: Pilot-Matias, Tami J
APPLICANT: Hunt, Jeffrey C
TITLE OF INVENTION: Borrelia burgdorferi antigens
and uses thereof
NUMBER OF SEQUENCES: 16

CORRESPONDENCE ADDRESS:

ADDRESSEE: Abbott Laboratories
STREET: One Abbott Park Road
CITY: Abbott Park
STATE: Illinois
COUNTRY: USA
ZIP: 60064

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/500,125
FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/779,704

FILING DATE: 21-OCT-1991

ATTORNEY/AGENT INFORMATION:

NAME: Wong, Wean Khing
REGISTRATION NUMBER: 33561
REFERENCE/DOCKET NUMBER: 5051.US.01

TELECOMMUNICATION INFORMATION:

TELEPHONE: 708-937-9396
TELEFAX: 708-937-9556

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 498 amino acids
TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein
HYPOTHETICAL: NO

ANTI-SENSE: NO

FRAGMENT TYPE: internal

ORIGINAL SOURCE:

ORGANISM: Borrelia burgdorferi
STRAIN: B31

US-08-500-125-2

Query Match

Best Local Similarity 1.2%; Score 7; DB 1; Length 498;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 540 SLGSQA 546

Db 350 SLGSQA 356

RESULT 26

US-07-779-704B-2

; Sequence 2, Application US/07779704B
; Patent No. 5965702

GENERAL INFORMATION:

APPLICANT: Robinson, John M
APPLICANT: Pilot-Matias, Tami J
APPLICANT: Hunt, Jeffrey C

; TITLE OF INVENTION: Borrelia burgdorferi antigens
and uses thereof
NUMBER OF SEQUENCES: 16

CORRESPONDENCE ADDRESS:

ADDRESSEE: Abbott Laboratories
STREET: One Abbott Park Road
CITY: Abbott Park
STATE: Illinois
COUNTRY: USA
ZIP: 60064

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/779,704B
FILING DATE: 19911021

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Wong, Wean Khing
REGISTRATION NUMBER: 33561
REFERENCE/DOCKET NUMBER: 5051.US.01

TELECOMMUNICATION INFORMATION:

TELEPHONE: 708-937-9396
TELEFAX: 708-937-9556

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 498 amino acids
TYPE: AMINO ACID

TOPOLOGY: linear

MOLECULE TYPE: protein
HYPOTHETICAL: NO

ANTI-SENSE: NO

FRAGMENT TYPE: internal

ORIGINAL SOURCE:

ORGANISM: Borrelia burgdorferi
STRAIN: B31

US-07-779-704B-2

Query Match

Best Local Similarity 1.2%; Score 7; DB 2; Length 498;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 540 SLGSQA 546

Db 350 SLGSQA 356

RESULT 27

US-08-796-899-25

; Sequence 25, Application US/08796899

; Patent No. 6160202

GENERAL INFORMATION:

APPLICANT: BUSTOS, Mauricio M

APPLICANT: CHERN, Maw-Sheng

TITLE OF INVENTION: MODIFICATION OF SEED CROPS WITH

TITLE OF INVENTION: TRANSCRIPTION FACTORS

NUMBER OF SEQUENCES: 32

CORRESPONDENCE ADDRESS:

ADDRESSEE: Burns, Doane, Swecker & Mathis

STREET: P.O. Box 1404

CITY: Alexandria

STATE: Virginia

COUNTRY: United States

ZIP: 22313-1404

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/796,899
FILING DATE: 06-FEB-1997
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/319,544
FILING DATE: 07-OCT-1994
ATTORNEY/AGENT INFORMATION:
NAME: Meuth, Donna M
REGISTRATION NUMBER: 36,607
REFERENCE/DOCKET NUMBER: 028754-005
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-6620
INFORMATION FOR SEQ ID NO: 25:
SEQUENCE CHARACTERISTICS:
LENGTH: 514 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-796-899-25

Query Match 1.2%; Score 7; DB 4; Length 514;
Best Local Similarity 100.0%; Pred. No. 89;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 147 SIKKELE 153
|||||||
Db 483 SIKKELE 489

RESULT 28
US-09-188-579-62
Sequence 62, Application US/09188579B
Patent No. 6107040
GENERAL INFORMATION:
APPLICANT: Shuman, Stewart
TITLE OF INVENTION: Pharmacological Targeting of mRNA Cap Formation
FILE REFERENCE: D6185
CURRENT APPLICATION NUMBER: US/09/188,579B
CURRENT FILING DATE: 1998-11-09
NUMBER OF SEQ ID NOS: 114
SEQ ID NO 62
LENGTH: 8
TYPE: PRT
ORGANISM: vaccinia virus
FEATURE:
OTHER INFORMATION: Motif III of RNA guanylyltransferase.
US-09-188-579-62

Query Match 1.0%; Score 6; DB 3; Length 8;
Best Local Similarity 100.0%; Pred. No. 1.4e+05;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 272 VVVFGE 277
|||||||
Db 1 VVVFGE 6

RESULT 29
US-08-960-780-69
Sequence 69, Application US/08960780
Patent No. 6204435
GENERAL INFORMATION:
APPLICANT: Feitelson, Jerald S.
APPLICANT: Schnepf, H. Ernest
APPLICANT: Narva, Kenneth E.
APPLICANT: Stockhoff, Brian A.
APPLICANT: Schmeits, James
APPLICANT: Loewer, David
APPLICANT: Dullum, Charles Joseph

APPLICANT: Muller-Cohn, Judy
APPLICANT: Stamp, Lisa
TITLE OF INVENTION: No. 6204435el Pesticidal Toxins and Nucleotide
TITLE OF INVENTION: Sequences Which Encode These Toxins
NUMBER OF SEQUENCES: 134
CORRESPONDENCE ADDRESS:
ADDRESSEE: Saliwanchik, Lloyd & Saliwanchik
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: FL
COUNTRY: US
ZIP: 32606-6669
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/960,780
FILING DATE: 30-OCT-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/029,848
FILING DATE: 30-OCT-1996
ATTORNEY/AGENT INFORMATION:
NAME: Saliwanchik, David R.
REGISTRATION NUMBER: 31,794
REFERENCE/DOCKET NUMBER: MA-708
TELECOMMUNICATION INFORMATION:
TELEPHONE: 352-375-8100
TELEFAX: 352-372-5800
INFORMATION FOR SEQ ID NO: 69:
SEQUENCE CHARACTERISTICS:
LENGTH: 8 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-960-780-69

Query Match 1.0%; Score 6; DB 4; Length 8;
Best Local Similarity 100.0%; Pred. No. 1.4e+05;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 398 KGRVAV 403
|||||||
Db 2 KGRVAV 7

RESULT 30
US-09-315-444-62
Sequence 62, Application US/09315444A
Patent No. 6232070
GENERAL INFORMATION:
APPLICANT: Shuman, Stewart
TITLE OF INVENTION: Pharmacological Targeting of mRNA Cap Formation
FILE REFERENCE: D6185CIP
CURRENT APPLICATION NUMBER: US/09/315,444A
CURRENT FILING DATE: 1999-05-20
PRIOR APPLICATION NUMBER: US 09/188,579
PRIOR FILING DATE: 1998-11-09
NUMBER OF SEQ ID NOS: 116
SEQ ID NO 62
LENGTH: 8
TYPE: PRT
ORGANISM: vaccinia virus
FEATURE:
OTHER INFORMATION: Motif III of RNA guanylyltransferase.
US-09-315-444-62

Query Match 1.0%; Score 6; DB 4; Length 8;

Best Local Similarity 100.0%; Pred. No. 1.4e+05;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 272 VVVEGE 277
Db 1 VVVEGE 6

RESULT 31

US-09-073-898-69
; Sequence 69, Application US/09073898
; Patent No. 6242669
; GENERAL INFORMATION:
; APPLICANT: Feitelson, Gerald S.
; APPLICANT: Schnepf, H. Ernest
; APPLICANT: Narva, Kenneth E.
; APPLICANT: Stockhoff, Brian A.
; APPLICANT: Schmelts, James
; APPLICANT: Loewer, David
; APPLICANT: Dullum, Charles Joseph
; APPLICANT: Muller-Cohn, Judy
; APPLICANT: Stamp, Lisa
; APPLICANT: Morrill, George
; APPLICANT: Finstad-Lee, Stacey
; TITLE OF INVENTION: No. 6242669el Pesticidal Toxins and Nucleotide
; TITLE OF INVENTION: Sequences Which Encode These Toxins
; NUMBER OF SEQUENCES: 144
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Saliwanchik, Lloyd & Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: US
; ZIP: 32606-6669
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/073,898
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/029,848
; FILING DATE: 30-OCT-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/960,780
; FILING DATE: 30-OCT-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Sanders, Jay M.
; REGISTRATION NUMBER: 39,355
; REFERENCE/DOCKET NUMBER: MA-708C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 352-375-8100
; TELEFAX: 352-372-5800
; INFORMATION FOR SEQ ID NO: 69:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-073-898-69

Query Match 1.0%; Score 6; DB 4; Length 8;
Best Local Similarity 100.0%; Pred. No. 1.4e+05;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 398 KGRVAV 403
Db 2 KGRVAV 7

RESULT 32
US-08-159-339A-1037
; Sequence 1037, Application US/08159339A
; Patent No. 6037135
; GENERAL INFORMATION:
; APPLICANT: Kubo, Ralph T.
; APPLICANT: Grey, Howard M.
; APPLICANT: Sette, Alessandro
; APPLICANT: Celis, Esteban
; TITLE OF INVENTION: HLA Binding peptides and Their
; TITLE OF INVENTION: Uses
; NUMBER OF SEQUENCES: 1254
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/159,339A
; FILING DATE: 29-NOV-1993
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/926,666
; FILING DATE: 07-AUG-1992
; APPLICATION NUMBER: US 08/027,746
; FILING DATE: 05-MAR-1993
; APPLICATION NUMBER: US 08/103,396
; FILING DATE: 06-AUG-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Weber, Ellen Lauver
; REGISTRATION NUMBER: 32,762
; REFERENCE/DOCKET NUMBER: 018623-005030US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; TELEX:
; INFORMATION FOR SEQ ID NO: 1037:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-159-339A-1037

Query Match 1.0%; Score 6; DB 3; Length 10;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 155 LVKEKG 160
Db 4 LVKEKG 9

RESULT 33

US-08-480-190-162
; Sequence 162, Application US/08480190
; Patent No. 5827516
; GENERAL INFORMATION:
; APPLICANT: Robert G. Urban
; APPLICANT: Roman M. Chicz
; APPLICANT: Dario A. A. Vignali
; APPLICANT: Mary L. Hedley

APPLICANT: Lawrence J. Stern
APPLICANT: Jack L. Strominger
TITLE OF INVENTION: IMMUNOMODULATORY PEPTIDES
NUMBER OF SEQUENCES: 274
CORRESPONDENCE ADDRESS:
STREET: Fish & Richardson
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
COMPUTER: IBM PS/2 Model 502 or 55SX
OPERATING SYSTEM: MS-DOS (Version 5.0)
SOFTWARE: WordPerfect (Version 5.1)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/480,190
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/077,255
FILING DATE: June 15, 1993
APPLICATION NUMBER: 07/925,460
FILING DATE: August 11, 1992
ATTORNEY/AGENT INFORMATION:
NAME: Clark, Paul T.
REGISTRATION NUMBER: 30,162
REFERENCE/DOCKET NUMBER: 00246/168001
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 162:
SEQUENCE CHARACTERISTICS:
LENGTH: 15
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-08-480-190-162

Query Match 1.0%; Score 6; DB 2; Length 15;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 236 VYRAVT 241
Db 4 VYRAVT 9

RESULT 34
US-08-488-379-162
Sequence 162, Application US/08488379
Patent No. 5860103
GENERAL INFORMATION:
APPLICANT: Robert G. Urban
APPLICANT: Roman M. Chicz
APPLICANT: Dario A. A. Vignali
APPLICANT: Mary L. Hedley
APPLICANT: Lawrence J. Stern
APPLICANT: Jack L. Strominger
TITLE OF INVENTION: IMMUNOMODULATORY PEPTIDES
NUMBER OF SEQUENCES: 274
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson
STREET: 225 Franklin Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb

COMPUTER: IBM PS/2 Model 502 or 55SX
OPERATING SYSTEM: MS-DOS (Version 5.0)
SOFTWARE: WordPerfect (Version 5.1)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/488,379
FILING DATE:
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/077,255
FILING DATE: June 15, 1993
APPLICATION NUMBER: 07/925,460
FILING DATE: August 11, 1992
ATTORNEY/AGENT INFORMATION:
NAME: Clark, Paul T.
REGISTRATION NUMBER: 30,162
REFERENCE/DOCKET NUMBER: 00246/168001
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 162:
SEQUENCE CHARACTERISTICS:
LENGTH: 15
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-08-488-379-162

Query Match 1.0%; Score 6; DB 2; Length 15;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 236 VYRAVT 241
Db 4 VYRAVT 9

RESULT 35
PCT-US93-07545-162
Sequence 162, Application PC/TUS9307545
GENERAL INFORMATION:
APPLICANT: Robert G. Urban
APPLICANT: Roman M. Chicz
APPLICANT: Dario A. A. Vignali
APPLICANT: Mary L. Hedley
APPLICANT: Lawrence J. Stern
APPLICANT: Jack L. Strominger
TITLE OF INVENTION: IMMUNOMODULATORY PEPTIDES
NUMBER OF SEQUENCES: 273
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson
STREET: 225 Franklin Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
COMPUTER: IBM PS/2 Model 502 or 55SX
OPERATING SYSTEM: MS-DOS (Version 5.0)
SOFTWARE: WordPerfect (Version 5.1)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/07545
FILING DATE: 19930811
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/925,460
FILING DATE: August 11, 1992
ATTORNEY/AGENT INFORMATION:
NAME: Clark, Paul T.
REGISTRATION NUMBER: 30,162
REFERENCE/DOCKET NUMBER: 00246/168001

TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 162:
SEQUENCE CHARACTERISTICS:
LENGTH: 15
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
PCT-US93-07545-162

Query Match 1.0%; Score 6; DB 5; Length 15;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 236 VYRAVT 241
Db 4 VYRAVT 9

RESULT 36
US-08-480-190-161
; Sequence 161, Application US/08480190
; Patent No. 5827516
; GENERAL INFORMATION:
; APPLICANT: Robert G. Urban
; APPLICANT: Roman M. Chicz
; APPLICANT: Dario A. A. Vignali
; APPLICANT: Mary L. Hedley
; APPLICANT: Lawrence J. Stern
; APPLICANT: Jack L. Strominger
; TITLE OF INVENTION: IMMUNOMODULATORY PEPTIDES
; NUMBER OF SEQUENCES: 274
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 502 or 55SX
; OPERATING SYSTEM: MS-DOS (Version 5.0)
; SOFTWARE: WordPerfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/480,190
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/077,255
; FILING DATE: June 15, 1993
; APPLICATION NUMBER: 07/925,460
; FILING DATE: August 11, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Clark, Paul T.
; REGISTRATION NUMBER: 30,162
; REFERENCE/DOCKET NUMBER: 00246/168001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-8906
; TELEFAX: 200154
; INFORMATION FOR SEQ ID NO: 161:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 17
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
US-08-480-190-161

QY 236 VYRAVT 241
Db 4 VYRAVT 9

US-08-480-190-161
; Sequence 161, Application US/08480190
; Patent No. 5827516
; GENERAL INFORMATION:
; APPLICANT: Robert G. Urban
; APPLICANT: Roman M. Chicz
; APPLICANT: Dario A. A. Vignali
; APPLICANT: Mary L. Hedley
; APPLICANT: Lawrence J. Stern
; APPLICANT: Jack L. Strominger
; TITLE OF INVENTION: IMMUNOMODULATORY PEPTIDES
; NUMBER OF SEQUENCES: 274
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 502 or 55SX
; OPERATING SYSTEM: MS-DOS (Version 5.0)
; SOFTWARE: WordPerfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/480,190
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/077,255
; FILING DATE: June 15, 1993
; APPLICATION NUMBER: 07/925,460
; FILING DATE: August 11, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Clark, Paul T.
; REGISTRATION NUMBER: 30,162
; REFERENCE/DOCKET NUMBER: 00246/168001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-8906
; TELEFAX: 200154
; INFORMATION FOR SEQ ID NO: 161:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 17
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
US-08-480-190-161

Query Match 1.0%; Score 6; DB 2; Length 17;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 236 VYRAVT 241
Db 4 VYRAVT 9

RESULT 37
US-08-488-379-161
; Sequence 161, Application US/08488379
; Patent No. 5880103
; GENERAL INFORMATION:
; APPLICANT: Robert G. Urban
; APPLICANT: Roman M. Chicz
; APPLICANT: Dario A. A. Vignali
; APPLICANT: Mary L. Hedley
; APPLICANT: Lawrence J. Stern
; APPLICANT: Jack L. Strominger
; TITLE OF INVENTION: IMMUNOMODULATORY PEPTIDES
; NUMBER OF SEQUENCES: 274
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 502 or 55SX
; OPERATING SYSTEM: MS-DOS (Version 5.0)
; SOFTWARE: WordPerfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/488,379
; FILING DATE:
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/077,255
; FILING DATE: June 15, 1993
; APPLICATION NUMBER: 07/925,460
; FILING DATE: August 11, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Clark, Paul T.
; REGISTRATION NUMBER: 30,162
; REFERENCE/DOCKET NUMBER: 00246/168001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 161:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 17
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
US-08-488-379-161

Query Match 1.0%; Score 6; DB 2; Length 17;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 236 VYRAVT 241
Db 4 VYRAVT 9

RESULT 38
PCT-US93-07545-161
; Sequence 161, Application PC/TUS9307545
; GENERAL INFORMATION:

APPLICANT: Robert G. Urban
APPLICANT: Roman M. Chicz
APPLICANT: Dario A. A. Vignali
APPLICANT: Mary L. Hedley
APPLICANT: Lawrence J. Stern
APPLICANT: Jack L. Strominger
TITLE OF INVENTION: IMMUNOMODULATORY PEPTIDES
NUMBER OF SEQUENCES: 273
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson
STREET: 225 Franklin Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
COMPUTER: IBM PS/2 Model 502 or 55SX
OPERATING SYSTEM: MS-DOS (Version 5.0)
SOFTWARE: WordPerfect (Version 5.1)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/07545
FILING DATE: 19930811
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/925,460
FILING DATE: August 11, 1992
ATTORNEY/AGENT INFORMATION:
NAME: Clark, Paul T.
REGISTRATION NUMBER: 30,162
REFERENCE/DOCKET NUMBER: 00246/168001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 161:
SEQUENCE CHARACTERISTICS:
LENGTH: 17
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
PCT-US93-07545-161

Query Match 1.0%; Score 6; DB 5; Length 17;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 236 VYRAVT 241
Db 4 VYRAVT 9

RESULT 39
PCT-US95-08596-21
Sequence 21, Application PC/TUS9508596
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: Proinsulin Peptide Compounds for Detecting
and Treating Type I Diabetes
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD
STREET: 60 State Street, suite 510
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109-1875
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/08596
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/272,220
FILING DATE: 08-JULY-1994
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: DeConti, Giulio A., Jr.
REGISTRATION NUMBER: 31,503
REFERENCE/DOCKET NUMBER: DCI-092PC
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)227-5941
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 17 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
FRAGMENT TYPE: internal
PCT-US95-08596-21

Query Match 1.0%; Score 6; DB 5; Length 17;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 6; Conservative .0; Mismatches 0; Indels 0; Gaps 0;

QY 155 LVKEKG 160
Db 8 LVKEKG 13

RESULT 40
PCT-US95-08621-20
Sequence 20, Application PC/TUS9508621
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: Invariant Chain Compositions Useful for
Treating Autoimmune Diseases
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD
STREET: 60 State Street, suite 510
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109-1875
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/08621
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/272,220
FILING DATE: 08-JULY-1994
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: DeConti, Giulio A., Jr.
REGISTRATION NUMBER: 31,503
REFERENCE/DOCKET NUMBER: DCI-070PC
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)227-5941
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 17 amino acids
TYPE: amino acid

TOPOLOGY: linear
MOLECULE TYPE: peptide
FRAGMENT TYPE: internal
PCT-US95-08621-20

Query Match 1.0%; Score 6; DB 5; Length 17;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 155 LVKEG 160
|||||
Db 8 LVKEG 13

RESULT 41
US-08-480-190-160
; Sequence 160, Application US/08480190
; Patent No. 5827516
; GENERAL INFORMATION:
; APPLICANT: Robert G. Urban
; APPLICANT: Roman M. Chicz
; APPLICANT: Dario A. A. Vignali
; APPLICANT: Mary L. Hedley
; APPLICANT: Lawrence J. Stern
; APPLICANT: Jack L. Strominger
; TITLE OF INVENTION: IMMUNOMODULATORY PEPTIDES
; NUMBER OF SEQUENCES: 274
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 502 or 55SX
; OPERATING SYSTEM: MS-DOS (Version 5.0)
; SOFTWARE: WordPerfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/480,190
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/077,255
; FILING DATE: June 15, 1993
; APPLICATION NUMBER: 07/925,460
; FILING DATE: August 11, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Clark, Paul T.
; REGISTRATION NUMBER: 30,162
; REFERENCE/DOCKET NUMBER: 00246/168001
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 160:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 18
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
US-08-480-190-160

Query Match 1.0%; Score 6; DB 2; Length 18;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 236 VYRAVT 241
|||||
Db 5 VYRAVT 10

RESULT 42
US-08-488-379-160
; Sequence 160, Application US/08488379
; Patent No. 5880103
; GENERAL INFORMATION:
; APPLICANT: Robert G. Urban
; APPLICANT: Roman M. Chicz
; APPLICANT: Dario A. A. Vignali
; APPLICANT: Mary L. Hedley
; APPLICANT: Lawrence J. Stern
; APPLICANT: Jack L. Strominger
; TITLE OF INVENTION: IMMUNOMODULATORY PEPTIDES
; NUMBER OF SEQUENCES: 274
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 502 or 55SX
; OPERATING SYSTEM: MS-DOS (Version 5.0)
; SOFTWARE: WordPerfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/488,379
; FILING DATE:
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/077,255
; FILING DATE: June 15, 1993
; APPLICATION NUMBER: 07/925,460
; FILING DATE: August 11, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Clark, Paul T.
; REGISTRATION NUMBER: 30,162
; REFERENCE/DOCKET NUMBER: 00246/168001
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 160:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 18
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
US-08-488-379-160

Query Match 1.0%; Score 6; DB 2; Length 18;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 236 VYRAVT 241
|||||
Db 5 VYRAVT 10

RESULT 43
PCT-US93-07545-160
; Sequence 160, Application PC/TUS9307545
; GENERAL INFORMATION:
; APPLICANT: Robert G. Urban
; APPLICANT: Roman M. Chicz
; APPLICANT: Dario A. A. Vignali
; APPLICANT: Mary L. Hedley
; APPLICANT: Lawrence J. Stern
; APPLICANT: Jack L. Strominger
; TITLE OF INVENTION: IMMUNOMODULATORY PEPTIDES

; NUMBER OF SEQUENCES: 273
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 502 or 55SX
; OPERATING SYSTEM: MS-DOS (Version 5.0)
; SOFTWARE: WordPerfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/07545
; FILING DATE: 19930811
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/925,460
; FILING DATE: August 11, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Clark, Paul T.
; REGISTRATION NUMBER: 30,162
; REFERENCE/DOCKET NUMBER: 00246/168001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-8906
; TELEFAX: (617) 542-5070
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 160:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 18
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
PCT-US93-07545-160

Query Match 1.0%; Score 6; DB 5; Length 18;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 236 VYRAVT 241
DB 5 VYRAVT 10

RESULT 44
US-08-468-543-10
; Sequence 10, Application US/08468543
; Patent No. 5726153
; GENERAL INFORMATION:
; APPLICANT: Lees, Robert S. et al.
; TITLE OF INVENTION: SYNTHETIC PEPTIDES FOR ARTERIAL IMAGING
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/468,543
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/398,046
; FILING DATE: 02-MAR-1995

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/201,057
; FILING DATE: 24-FEB-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/048,569
; FILING DATE: 16-APR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/694,929
; FILING DATE: 02-MAY-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/518,215
; FILING DATE: 03-MAY-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/518,142
; FILING DATE: 03-MAY-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/189,130
; FILING DATE: 02-MAY-1988
; ATTORNEY/AGENT INFORMATION:
; NAME: Clark, Paul T.
; REGISTRATION NUMBER: 30,162
; REFERENCE/DOCKET NUMBER: 04547/002003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
US-08-468-543-10

Query Match 1.0%; Score 6; DB 1; Length 20;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 150 RELEAL 155
DB 12 RELEAL 17

RESULT 45
US-08-469-692-10
; Sequence 10, Application US/08469692
; Patent No. 5955055
; GENERAL INFORMATION:
; APPLICANT: Lees, Robert S. et al.
; TITLE OF INVENTION: SYNTHETIC PEPTIDES FOR ARTERIAL IMAGING
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/469,692
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/398,046
; FILING DATE: 02-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/201,057

;; FILING DATE: 24-FEB-1994
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/048,569
;; FILING DATE: 16-APR-1993
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/694,929
;; FILING DATE: 02-MAY-1991
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/518,215
;; FILING DATE: 03-MAY-1990
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/518,142
;; FILING DATE: 03-MAY-1990
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/189,130
;; FILING DATE: 02-MAY-1988
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Clark, Paul T.
;; REGISTRATION NUMBER: 30,162
;; REFERENCE/DOCKET NUMBER: 04547/002002
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 617/542-5070
;; TELEFAX: 617/542-8906
;; TELEX: 200154
;; INFORMATION FOR SEQ ID NO: 10:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 20
;; TYPE: amino acid
;; STRANDEDNESS:
;; TOPOLOGY: linear
;; US-08-469-692-10

Query Match 1.0%; Score 6; DB 2; Length 20;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 150 EELEAL 155
Db 12 BELEAL 17

Search completed: July 30, 2001, 11:43:18
Job time: 137 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2001, 11:41:26 ; Search time 15.34 Seconds
(without alignments)
2840.407 Million cell updates/sec

Title: US-09-367-496-8
Perfect score: 572
Sequence: 1 MSFQKKSIPRTSDRLIR.....RTAQKIWPPGGRSNTLSLS 572

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 219241 seqs, 76174552 residues

Word size : 0

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : PIR_68:*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	31	5.4	572	2 J49985	dihydropyrimidinase
2	31	5.4	572	2 J5317	dihydropyrimidinase
3	31	5.4	572	2 S5889	collapsin response
4	18	3.1	570	2 J5318	dihydropyrimidinase
5	18	3.1	570	2 J5318	Ulip protein - mou
6	18	3.1	572	2 J5316	dihydropyrimidinase
7	14	2.4	519	2 J5315	dihydropyrimidinase
8	14	2.4	519	2 S70581	dihydropyrimidinase
9	11	1.9	854	2 S3358	unc-33 protein - C
10	8	1.4	309	1 B64041	hypothetical prote
11	8	1.4	376	2 T51673	myb-related transc
12	8	1.4	465	2 A65071	hypothetical prote
13	8	1.4	465	2 F85942	hypothetical prote
14	8	1.4	520	2 T20007	hypothetical prote
15	8	1.4	545	2 A32260	cholesterol oxidas
16	8	1.4	648	2 T47988	serine/threonine-p
17	8	1.4	900	2 E83189	protein-PII uridy
18	7	1.2	10	2 G60787	sperm-activating p
19	7	1.2	10	2 E60787	sperm-activating p
20	7	1.2	76	2 T51499	hypothetical prote
21	7	1.2	89	2 J00387	hypothetical prote
22	7	1.2	94	2 C69484	conserved hypothet
23	7	1.2	117	1 GARS2K	Ig heavy chain V-A
24	7	1.2	119	1 G1HUDW	Ig heavy chain V-I
25	7	1.2	123	2 H83903	hypothetical prote
26	7	1.2	124	2 S73210	ribosomal protein
27	7	1.2	127	2 C81292	hypothetical prote
28	7	1.2	135	2 C75465	conserved hypothet
29	7	1.2	138	2 C70077	hypothetical prote

conserved hypothet
RCc2 protein - ric
RCg2 protein - ric
hypothetical prote
hypothetical prote
type 4 fibrillar bi
hypothetical prote
probable lipoprote
hypothetical prote
hypothetical prote
troponin I - fruit
troponin I - fruit
hypothetical prote
p-nitrophenyl phos
sex-determining pr
SOS response regul
hypothetical prote

11.2 141 2 D69857
1.2 146 2 S53010
1.2 146 2 S53011
1.2 147 2 T16440
1.2 147 2 T30616
1.2 185 2 S77594
1.2 191 2 S76398
1.2 197 2 F81934
1.2 205 2 T33826
1.2 208 2 A38594
1.2 208 2 A40547
1.2 210 2 E72385
1.2 212 2 H84244
1.2 227 2 S35564
1.2 234 2 T35123
1.2 235 2 F70405

ALIGNMENTS

RESULT 1
S49985
dihydropyrimidinase-related protein 2 [similarity] - rat
N:Alternate names: collapsin response mediator protein 2; TOND-64
C:Species: Rattus norvegicus (Norway rat)
C:Date: 13-Jan-1995 #sequence_revision 12-May-1995 #text_change 02-Sep-2000
C:Accession: A59280; S49985
R:Minturn, J.E.; Fryer, H.J.; Geschwind, D.H.; Hockfield, S.
J. Neurosci. 15, 6757-6766, 1995
A:Title: TOND-64, a gene expressed early in neuronal differentiation in the rat, is r
A:Reference number: A59280; MUID:96033765
A:Accession: A59280
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-572 <MIN>
A:Cross-references: GB:246882; NID:g599965; PIDN:CAA86981.1; PID:g599966
A:Experimental source: strain Sprague-Dawley; clone TOND64 (turned on after division,
A:Note: submitted to the EMBL Data Library, November 1994
A:Note: in Genbank entry RRTOND64, release 113.0, the source is designated as Rattus
C:Superfamily: allantoinase; Bacillus dihydroorotase homology

Query Match 5.4%; Score 31; DB 2; Length 572;
Best Local Similarity 100.0%; Pred. No. 3.7e-23;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 276 GEPITASLGTDGSHYWSKNWAKAAAFVTSP 306
|||||
DB 276 GEPITASLGTDGSHYWSKNWAKAAAFVTSP 306

RESULT 2
J5317
dihydropyrimidinase-related protein 2 - human
N:Alternate names: collapsin response mediator protein 2
C:Species: Homo sapiens (man)
C:Date: 01-May-1997 #sequence_revision 23-Aug-1997 #text_change 02-Sep-2000
C:Accession: J5317; S58891; J53150
R:HamaJima, N.; Matsuda, K.; Sakata, S.; Tamaki, M.; Nonaka, M.
Gene 180, 157-163, 1996
A:Title: A novel gene family defined by human dihydropyrimidinase and three related p
A:Reference number: J5317; MUID:97128821
A:Accession: J5317
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-572 <HAM>
A:Cross-references: DDBJ:D78013; NID:g1330239; PIDN:BA11191.1; PID:g1330240
A:Experimental source: fetal brain
R:Goshima, Y.; Nakamura, F.; Strittmatter, P.; Strittmatter, S.M.
Nature 376, 509-514, 1995
A:Title: Collapsin-induced growth cone collapse mediated by an intracellular protein
A:Reference number: S5889; MUID:95364923

A:Accession: S58891
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: mRNA
 A:Residues: 1-572 <GOS>
 A:Cross-references: EMBL:U17279; NID:gl244399; PIDN:AAA93202.1; PID:gl244400
 A:Note: the nucleotide sequence was submitted to the EMBL data library, April 1996
 R:Kitamura, K.; Takayama, M.; Hamajima, N.; Nakanishi, M.; Sasaki, M.; Endo, Y.; Takemoto
 DNA Res. 6, 291-297, 1999
 A:Title: Characterization of the human dihydropyrimidinase-related protein 2 (DRP-2) gene
 A:Reference number: JC7150; MUID:20039612
 A:Accession: JC7150
 A:Status: preliminary
 A:Molecule type: nucleic acid
 A:Residues: 1-572 <KIT>
 A:Cross-references: DBJ:AB020764; PIDN:BA89422.1
 A:Comment: This protein is the homologue of chicken 62K collapsin response mediator protein
 C:Genetics:
 A:Gene: GDB:DPYSL2; DRP2; CRMP2
 A:Cross-references: GDB:5496401
 A:Map position: 8p22-8p21
 C:Superfamily: allantoinase; Bacillus dihydroorotase homology

Query Match 5.4%; Score 31; DB 2; Length 572;
 Best Local Similarity 100.0%; Pred. No. 3.7e-23;
 Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 276 GEPITASLGTGSHYWSKNWAKAAAFVTSPP 306
 |||||
 DB 276 GEPITASLGTGSHYWSKNWAKAAAFVTSPP 306

RESULT 3
 S58889
 collapsin response mediator protein, 62K - chicken
 C:Species: Gallus gallus (chicken)
 C:Date: 15-Feb-1996 #sequence_revision 01-Mar-1996 #text_change 02-Sep-2000
 C:Accession: S58889
 R:Goshima, Y.; Nakamura, F.; Strittmatter, P.; Strittmatter, S.M.
 Nature 376, 509-514, 1995
 A:Title: Collapsin-induced growth cone collapse mediated by an intracellular protein rel
 A:Reference number: S58889; MUID:95364923
 A:Accession: S58889
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-572 <GOS>
 A:Cross-references: EMBL:U17277; NID:g882146; PIDN:AAA93200.1; PID:g882147
 C:Superfamily: allantoinase; Bacillus dihydroorotase homology

Query Match 5.4%; Score 31; DB 2; Length 572;
 Best Local Similarity 100.0%; Pred. No. 3.7e-23;
 Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 276 GEPITASLGTGSHYWSKNWAKAAAFVTSPP 306
 |||||
 DB 276 GEPITASLGTGSHYWSKNWAKAAAFVTSPP 306

RESULT 4
 JC5318
 dihydropyrimidinase related protein 3 - human
 C:Species: Homo sapiens (man)
 C:Date: 01-May-1997 #sequence_revision 18-Jul-1997 #text_change 02-Sep-2000
 C:Accession: JC5318
 R:Hamajima, N.; Matsuda, K.; Sakata, S.; Tamaki, N.; Sasaki, M.; Nonaka, M.
 Gene 180, 157-163, 1996
 A:Title: A novel gene family defined by human dihydropyrimidinase and three related prot
 A:Reference number: JC5315; MUID:97128821
 A:Accession: JC5318
 A:Status: nucleic acid sequence not shown
 A:Molecule type: mRNA
 A:Residues: 1-570 <HAM>

A:Cross-references: DBJ:D78014; NID:gl330241; PIDN:BAAL1192.1; PID:gl330242
 A:Experimental source: fetal brain
 C:Comment: This protein is the homologue of rat 64K turned on after division protein
 C:Genetics:
 A:Gene: GDB:DPYSL3; DRP-3
 A:Cross-references: GDB:5885804
 C:Superfamily: allantoinase; Bacillus dihydroorotase homology

Query Match 3.1%; Score 18; DB 2; Length 570;
 Best Local Similarity 100.0%; Pred. No. 8.2e-10;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 289 HYWSKNWAKAAAFVTSPP 306
 |||||
 DB 289 HYWSKNWAKAAAFVTSPP 306

RESULT 5
 S55525
 Ulip protein - mouse
 C:Species: Mus musculus (house mouse)
 C:Date: 01-Aug-1995 #sequence_revision 01-Sep-1995 #text_change 02-Sep-2000
 C:Accession: S55525
 R:Byk, T.; Dobransky, T.; Cifuentes-Diaz, C.; Sobel, A.
 submitted to the EMBL data library, June 1995
 A:Description: Identification and molecular characterisation of Ulip, a putative mamm
 A:Reference number: S55525
 A:Accession: S55525
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-570 <BYK>
 A:Cross-references: EMBL:X87817; NID:g861058; PIDN:CAA61082.1; PID:g861059
 C:Genetics:
 A:Gene: Ulip
 C:Superfamily: allantoinase; Bacillus dihydroorotase homology

Query Match 3.1%; Score 18; DB 2; Length 570;
 Best Local Similarity 100.0%; Pred. No. 8.2e-10;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 289 HYWSKNWAKAAAFVTSPP 306
 |||||
 DB 289 HYWSKNWAKAAAFVTSPP 306

RESULT 6
 JC5316
 dihydropyrimidinase related protein 1 - human
 C:Species: Homo sapiens (man)
 C:Date: 01-May-1997 #sequence_revision 18-Jul-1997 #text_change 01-Dec-2000
 C:Accession: JC5316; S58890
 R:Hamajima, N.; Matsuda, K.; Sakata, S.; Tamaki, N.; Sasaki, M.; Nonaka, M.
 Gene 180, 157-163, 1996
 A:Title: A novel gene family defined by human dihydropyrimidinase and three related p
 A:Reference number: JC5315; MUID:97128821
 A:Accession: JC5316
 A:Status: nucleic acid sequence not shown
 A:Molecule type: mRNA
 A:Residues: 1-572 <HAM>
 A:Cross-references: DBJ:D78012; NID:gl330237; PIDN:BAAL1190.1; PID:gl330238
 A:Experimental source: fetal brain
 R:Goshima, Y.; Nakamura, F.; Strittmatter, P.; Strittmatter, S.M.
 Nature 376, 509-514, 1995
 A:Title: Collapsin-induced growth cone collapse mediated by an intracellular protein
 A:Reference number: S58889; MUID:95364923
 A:Accession: S58890
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 64-503, 'H', 505-572 <GOS>
 A:Cross-references: EMBL:U17278; NID:g882148; PIDN:AAA93201.1; PID:g882149
 A:Note: the nucleotide sequence was submitted to the EMBL data library, November 1994

C:Comment: This protein is involved in nervous system development.

C:Genetics:
A:Gene: GDB:CRMP1; DRP-1; DPYSL1
A:Cross-references: GDB:5585714
C:Superfamily: allantoinase; Bacillus dihydroorotase homology

Query Match 3.1%; Score 18; DB 2; Length 572;
Best Local Similarity 100.0%; Pred. No. 8.2e-10;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 289 HYWKNWAKAAAFVTSPP 306
|||||
Db 289 HYWKNWAKAAAFVTSPP 306

RESULT 7
JC5315
dihydropyrimidinase (EC 3.5.2.2) - human
N:Alternate names: 5,6-dihydropyrimidine amidohydrolase; Hydantoinase
C:Species: Homo sapiens (man)
C:Date: 01-May-1997 #sequence_revision 18-Jul-1997 #text_change 18-Aug-2000
C:Accession: JC5315
R:Hamaajima, N.; Matsuda, K.; Sakata, S.; Tamaki, N.; Nonaka, M.
A:Title: A novel gene family defined by human dihydropyrimidinase and three related proteins
A:Reference number: JC5315; MUID:97128821
A:Accession: JC5315
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-519 <HAM>
A:Cross-references: DDBJ:D78011; NID:g2339965; PIDN:BAAL1189.1; PID:g1330236
A:Experimental source: liver
C:Comment: This enzyme is Zn2+-metalloenzyme, and the second enzyme involved in uracil and
dihydropyrimidine to N-carbamyl-beta-aminoisobutyrate. It catalyzes the hydrolysis of a variety
C:Genetics:
A:Gene: GDB:DPYS; DHPase
A:Cross-references: GDB:5885803
C:Complex: homotetramer
C:Superfamily: allantoinase; Bacillus dihydroorotase homology
C:Keywords: hydrolase

Query Match 2.4%; Score 14; DB 2; Length 519;
Best Local Similarity 100.0%; Pred. No. 9.6e-06;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 380 FVAVTSTNAKIFN 393
|||||
Db 374 FVAVTSTNAKIFN 387

RESULT 8
S70581
dihydropyrimidinase - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 14-Feb-1997 #sequence_revision 13-Mar-1997 #text_change 18-Aug-2000
C:Accession: S70581
R:Matsuda, K.; Sakata, S.; Kaneko, M.; Hamaajima, N.; Nonaka, M.; Sasaki, M.; Tamaki, N.
Biochim. Biophys. Acta 1307, 140-144, 1996
A:Title: Molecular cloning and sequencing of a cDNA encoding dihydropyrimidinase from the
A:Reference number: S70581; MUID:96283806
A:Accession: S70581
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-519 <MAT>
A:Cross-references: EMBL:D63704; NID:g1378018; PIDN:BAA09833.1; PID:d1010479; PID:g13780
C:Superfamily: allantoinase; Bacillus dihydroorotase homology

Query Match 2.4%; Score 14; DB 2; Length 519;
Best Local Similarity 100.0%; Pred. No. 9.6e-06;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 91 QGKKAALAGGTTMI 104
|||||
Db 85 QGKKAALAGGTTMI 98

RESULT 9
S33558
unc-33 protein - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 07-Apr-1994 #sequence_revision 02-Jun-1994 #text_change 21-Jul-2000
A:Accession: S33558; S33559; S24643; S24644
R:Li, W.; Herman, R.K.; Shaw, J.E.
Genetics 132, 675-689, 1992
A:Title: Analysis of the Caenorhabditis elegans axonal guidance and outgrowth gene un
A:Reference number: S33558; MUID:93106371
A:Accession: S33558
A:Molecule type: DNA
A:Residues: 1-854 <LIW1>
A:Cross-references: EMBL:Z14148; NID:g6899; PIDN:CAA78520.1; PID:g6900
A:Accession: S33559
A:Molecule type: mRNA
A:Residues: 1-854 <LIW2>
A:Cross-references: EMBL:Z14146; NID:g6903; PID:g6904
C:Genetics:
A:Introns: 14/2; 68/1; 129/3; 151/3; 182/1; 255/3; 361/2; 660/1; 790/3
C:Keywords: alternative initiators
F:1-854/Product: unc-33 protein (long form) #status predicted <MAT1>
F:176-854/Product: unc-33 protein (intermediate form) #status predicted <MAT2>
F:332-854/Product: unc-33 protein (short form) #status predicted <MAT3>

Query Match 1.9%; Score 11; DB 2; Length 854;
Best Local Similarity 100.0%; Pred. No. 0.018;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 380 FVAVTSTNAK 390
|||||
Db 683 FVAVTSTNAK 693

RESULT 10
B64041
hypothetical protein H11730 - Haemophilus influenzae (strain Rd KW20)
C:Species: Haemophilus influenzae
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 28-Jul-2000
C:Accession: B64041
R:Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage
; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman
; D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghegan, N.S.M.
Science 269, 496-512, 1995
A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Vente
A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A:Reference number: A64000; MUID:95350630
A:Accession: B64041
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-309 <TRGR>
A:Cross-references: GB:U032845; GB:L42023; NID:g3212236; PIDN:AAC23376.1; PID:g1574586
C:Superfamily: hypothetical protein H11730

Query Match 1.4%; Score 8; DB 1; Length 309;
Best Local Similarity 100.0%; Pred. No. 8.7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 190 DLGALAQV 197
|||||
Db 264 DLGALAQV 271

RESULT 11
T51673

myb-related transcription factor MYB68 [imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 18-Aug-2000 #sequence_revision 18-Aug-2000 #text_change 02-Sep-2000
 C:Accession: T51673
 R:Kranz, H.D.; Denekamp, M.; Greco, R.; Jin, H.; Kranz, H.D.; Denekamp, M.; Greco, R.; J.
 Paz-Ares, J.; Weishaar, B.
 Plant J. 16, 263-276, 1998
 A:Title: Towards functional characterisation of the members of the R2R3-MYB gene from A.
 R:Reference number: 214349; MUID:9839469
 A:Accession: T51673
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-376 <KRA>
 A:Cross-references: EMBL:AF062901; PIDN:AAC83623.1
 A:Experimental source: cultivar Columbia
 C:Genetics:
 A:Gene: MYB68
 A:Map position: V
 C:Superfamily: Arabidopsis myb-related protein M4; myb DNA-binding repeat homology
 C:Keywords: transcription factor

Query Match 1.4%; Score 8; DB 2; Length 376;
 Best Local Similarity 100.0%; Pred. No. 10;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 95 AALAGGTT 102
 |||||
 Db 337 AALAGGTT 344

RESULT 12

A65071
 hypothetical protein b2873 - Escherichia coli (strain K-12)
 C:Species: Escherichia coli
 C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 22-Jun-1999
 C:Accession: A65071
 R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
 A.; Rose, D.J.; Mau, B.; Shao, Y.
 Science 277, 1453-1462, 1997
 A:Title: The complete genome sequence of Escherichia coli K-12.
 A:Reference number: A64720; MUID:97426617

A:Accession: A65071
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-465 <BLAT>
 A:Cross-references: GB:A6000370; GB:U00096; NID:g2367170; PIDN:AAC75911.1; PID:g1789237;
 A:Experimental source: strain K-12, substrain M61655
 C:Superfamily: allantoinase; Bacillus dihydroorotase homology
 F:49-452/Domain: Bacillus dihydroorotase homology <DHOS>

Query Match 1.4%; Score 8; DB 2; Length 465;
 Best Local Similarity 100.0%; Pred. No. 13;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 67 PGGVDPVHT 74
 |||||
 Db 57 PGGVDPVHT 64

RESULT 13

F85942
 hypothetical protein Z4212 [imported] - Escherichia coli (strain O157:H7)
 C:Species: Escherichia coli
 C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 31-Mar-2001
 C:Accession: F85942
 R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
 Miller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
 Nature 409, 529-533, 2001
 A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
 A:Reference number: A85480; MUID:21074935; PMID:11206551
 A:Accession: F85942

A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-465 <STO>
 A:Cross-references: GB:A6005174; NID:g12517394; PIDN:AAG58002.1; GSPDB:GN00145; UWGP:
 A:Experimental source: strain O157:H7, substrain EDL933
 C:Genetics:
 A:Gene: Z4212
 C:Superfamily: allantoinase; Bacillus dihydroorotase homology

Query Match 1.4%; Score 8; DB 2; Length 465;
 Best Local Similarity 100.0%; Pred. No. 13;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 67 PGGVDPVHT 74
 |||||
 Db 57 PGGVDPVHT 64

RESULT 14

T20007
 hypothetical protein C47E12.8 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Aug-2000
 C:Accession: T20007
 R:Coles, L.
 submitted to the EMBL Data Library, January 1996
 A:Reference number: Z19210
 A:Accession: T20007
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-520 <WIL>
 A:Cross-references: EMBL:Z68882; PIDN:CAA93104.1; GSPDB:GN00022; CESP:C47E12.8
 A:Experimental source: clone C47E12
 C:Genetics:
 A:Gene: CESP:C47E12.8
 A:Map position: 4
 A:Introns: 29/2; 89/3; 147/1; 232/3; 310/2; 414/3
 C:Superfamily: allantoinase; Bacillus dihydroorotase homology

Query Match 1.4%; Score 8; DB 2; Length 520;
 Best Local Similarity 100.0%; Pred. No. 14;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 216 GITGPEGH 223
 |||||
 Db 203 GITGPEGH 210

RESULT 15

A32260
 cholesterol oxidase (EC 1.1.3.6) precursor [validated] - Streptomyces sp.
 C:Species: Streptomyces sp.
 C:Date: 20-Oct-1989 #sequence_revision 20-Oct-2000 #text_change 20-Oct-2000
 C:Accession: A32260; S15810; PC2002
 R:Ishizaki, T.; Hirayama, N.; Shinkawa, H.; Niml, O.; Murooka, Y.
 J. Bacteriol. 171, 596-601, 1989
 A:Title: Nucleotide sequence of the gene for cholesterol oxidase from a Streptomyces
 A:Reference number: A32260; MUID:89123081
 A:Accession: A32260

A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 'V', 2-546 <ISH>
 A:Cross-references: GB:M31939; GB:J03356; NID:g153210; PIDN:AAA26719.1; PID:g153212
 A:Note: the authors translated the initiation codon GTG for residue 1 as Val
 R:Horii, M.; Ishizaki, T.; Paik, S.Y.; Manome, T.; Murooka, Y.
 J. Bacteriol. 172, 3644-3653, 1990
 A:Title: An operon containing the genes for cholesterol oxidase and a cytochrome P-45
 A:Reference number: S15809; MUID:90299781
 A:Accession: S15810
 A:Molecule type: DNA
 A:Residues: 1-30 <HOR>

A:Cross-references: EMBL:M31939; GB:J03356; NID:g153210
 R:Purcell, J.P.; Greenplate, J.T.; Jennings, M.G.; Kyser, J.C.; Pershing, J.C.; Sims, S.
 Biochem. Biophys. Res. Commun. 196, 1406-1413, 1993
 A:Title: Cholesterol oxidase: a potent insecticidal protein active against boll weevil
 A:Reference number: PC2002; MUID:94071904
 A:Accession: PC2002
 A:Molecule type: protein
 A:Residues: 'XXXTF', 48-54, 'XX', 57, 'X', 59-60 <PUR>
 C:Genetics:
 A:Gene: choA
 C:Keywords: oxidoreductase
 F:1-42/Domain: signal sequence #status predicted <SIG>
 F:43-546/Product: cholesterol oxidase #status experimental <MAT>

Query Match 1.4%; Score 8; DB 2; Length 546;
 Best Local Similarity 100.0%; Pred. No. 14;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 95 AALAGGTT 102
 |||||
 Db 21 AALAGGTT 28

RESULT 16
 T47988
 serine/threonine-protein kinase-like protein - Arabidopsis thaliana
 N:Alternate names: protein F21F14.130
 C:Species: Arabidopsis thaliana (mouse-ear cross)
 C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
 C:Accession: T47988
 R:Cholsne, N.; Robert, C.; Brottier, P.; Wincker, P.; Cattolico, L.; Artiguenave, F.; Sa
 submitted to the Protein Sequence Database, February 2000
 A:Reference number: Z24481
 A:Accession: T47988
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-648 <CHO>
 A:Cross-references: EMBL:AL138642
 A:Experimental source: cultivar Columbia; BAC clone F21F14
 C:Genetics:
 A:Map position: 3
 A:Introns: 77/3; 117/1; 136/3; 158/2; 189/3; 216/3; 254/1; 272/2; 483/3; 563/1; 586/3
 A:Note: F21F14.130

Query Match 1.4%; Score 8; DB 2; Length 648;
 Best Local Similarity 100.0%; Pred. No. 17;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 307 VNPDPPTA 314
 |||||
 Db 515 VNPDPPTA 522

RESULT 17
 E83189
 protein-PII uridylyltransferase PA3658 [imported] - Pseudomonas aeruginosa (strain PA01)
 C:Species: pseudomonas aeruginosa
 C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
 C:Accession: E83189
 R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Miziochuchi, S.D.; Warren, P.; Hickey, M.J.; Br
 adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
 ; Lory, S.; Olson, M.V.
 Nature 406, 959-964, 2000
 A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
 A:Reference number: A82950; MUID:20437337
 A:Accession: E83189
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-900 <STO>
 A:Cross-references: GB:AE004785; GB:AE004091; NID:g9949809; PIDN:AAG07046.1; GSPDB:GN001
 A:Experimental source: strain PA01

C:Genetics:
 A:Gene: glnD; PA3658
 C:Superfamily: uridylyltransferase

Query Match 1.4%; Score 8; DB 2; Length 900;
 Best Local Similarity 100.0%; Pred. No. 23;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 326 LQVTGSAH 333
 |||||
 Db 171 LQVTGSAH 178

RESULT 18
 G60787
 sperm-activating peptide (Ser-3,5,7 speract) - sea urchin (Anthocidaris crassispina)
 C:Species: Anthocidaris crassispina
 C:Date: 03-Jun-1993 #sequence_revision 03-Jun-1993 #text_change 18-Aug-2000
 C:Accession: G60787
 R:Suzuki, N.; Kajiuura, H.; Nomura, K.; Garbers, D.L.; Yoshino, K.; Kurita, M.; Tanaka
 Comp. Biochem. Physiol. B 89, 687-693, 1988
 A:Title: Some more speract derivatives associated with eggs of sea urchins, Pseudocen
 A:Reference number: A60787; MUID:88242184
 A:Accession: G60787
 A:Molecule type: protein
 A:Residues: 1-10 <SUZ>
 C:Comment: This oligopeptide from egg jelly is one of several from this species, all
 at shows some, but not absolute, species restriction.
 C:Superfamily: unassigned animal peptides

Query Match 1.2%; Score 7; DB 2; Length 10;
 Best Local Similarity 100.0%; Pred. No. 4.3;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 538 GFSLSGS 544
 |||||
 Db 1 GFSLSGS 7

RESULT 19
 E60787
 sperm-activating peptide (Ser-3,5,7, Asp-10 speract) - sea urchin (Hemicentrotus pulc
 C:Species: Hemicentrotus pulcherrimus
 C:Date: 03-Jun-1993 #sequence_revision 03-Jun-1993 #text_change 18-Aug-2000
 C:Accession: E60787
 R:Suzuki, N.; Kajiuura, H.; Nomura, K.; Garbers, D.L.; Yoshino, K.; Kurita, M.; Tanaka
 Comp. Biochem. Physiol. B 89, 687-693, 1988
 A:Title: Some more speract derivatives associated with eggs of sea urchins, Pseudocen
 A:Reference number: A60787; MUID:88242184
 A:Accession: E60787
 A:Molecule type: protein
 A:Residues: 1-10 <SUZ>
 C:Comment: This oligopeptide from egg jelly is one of several from this species, all
 at shows some, but not absolute, species restriction.
 C:Superfamily: unassigned animal peptides

Query Match 1.2%; Score 7; DB 2; Length 10;
 Best Local Similarity 100.0%; Pred. No. 4.3;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 538 GFSLSGS 544
 |||||
 Db 1 GFSLSGS 7

RESULT 20
 T51499
 hypothetical protein F5E19.10 - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cross)
 C:Date: 18-Aug-2000 #sequence_revision 18-Aug-2000 #text_change 18-Aug-2000

C:Accession: T51499
R:Sato, S.; Nakamura, Y.; Kaneko, T.; Kato, T.; Asamizu, E.; Kotani, H.; Tabata, S.; Mew
submitted to the Protein Sequence Database, August 2000
A:Reference number: 225394
A:Accession: T51499
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-76 <SAT>
A:Cross-references: EMBL:AL391147
A:Experimental source: cultivar Columbia; BAC clone F5E19
C:Genetics:
A:Map position: 5
A:Introns: 14/1; 60/2
A:Note: F5E19_10

Query Match 1.2%; Score 7; DB 2; Length 76;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 230 EVEAEAV 236
|||||
Db 62 EVEAEAV 68

RESULT 21
JU0387
hypothetical 9.3K protein - mouse
C:Species: Mus musculus (house mouse)
C:Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 05-Nov-1999
C:Accession: JU0387
R:Kanai, A.; Kuzuhara, T.; Sekimizu, K.; Natori, S.
J. Biochem. 109, 674-677, 1991
A:Title: Heterogeneity and tissue-specific expression of eukaryotic transcription factor
A:Reference number: JX0167; MUID:92011448
A:Molecule type: mRNA
A:Residues: 1-89 <KAN>
A:Cross-references: GB:D00925; NID:g220590; PIDN:BAA00767.1; PID:dl001229; PID:g220591

Query Match 1.2%; Score 7; DB 2; Length 89;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 449 QGRVALE 455
|||||
Db 66 QGRVALE 72

RESULT 22
C69484
conserved hypothetical protein AF1876 - Archaeoglobus fulgidus
C:Species: Archaeoglobus fulgidus
C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 22-Oct-1999
C:Accession: C69484
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson
; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.
; Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
Nature 390, 364-370, 1997
A:Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Artiach, P.; Kaine, B.P.; Sykes, S.
Smith, H.O.; Woese, C.R.; Venter, J.C.
A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeo
A:Reference number: A69250; MUID:98049343
A:Accession: C69484
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-94 <KLE>
A:Cross-references: GB:AE000973; GB:AE000782; NID:g2689296; PIDN:AAB89376.1; PID:g264866

Query Match 1.2%; Score 7; DB 2; Length 94;
Best Local Similarity 100.0%; Pred. No. 32;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 155 LVKEGV 161
|||||
Db 85 LVKEGV 91

RESULT 23
GARB2K
Ig heavy chain V-A2 region (K-25) - rabbit (tentative sequence)
C:Species: Oryctolagus cuniculus (domestic rabbit)
C:Date: 07-May-1981 #sequence_revision 07-May-1981 #text_change 31-Mar-2000
C:Accession: A02104
R:Jaton, J.C.
Biochem. J. 147, 235-247, 1975
A:Title: Comparison of the amino acid sequences of the variable domains of two homoge
A:Reference number: A90281; MUID:76039436
A:Accession: A02104
A:Molecule type: protein
A:Residues: 1-117 <JAT>
C:Comment: This chain was obtained from antibody to type III pneumococci and was isol
C:Superfamily: immunoglobulin V region; immunoglobulin homology
C:Keywords: blocked amino end; heterotetramer; immunoglobulin
F:14-93/Domain: immunoglobulin homology <IMM>
F:1/Modified site: blocked amino end (Gln) (probably pyrrolidone carboxylic acid) #st
F:21-91/Disulfide bonds: #status experimental

Query Match 1.2%; Score 7; DB 1; Length 117;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 537' SGFSLSG 543
|||||
Db 24 SGFSLSG 30

RESULT 24
GIHUDW
Ig heavy chain V-II region (Daw) - human
C:Species: Homo sapiens (man)
C:Date: 23-Oct-1981 #sequence_revision 23-Oct-1981 #text_change 02-Sep-1997
C:Accession: A02091
R:Press, E.M.; Hogg, N.M.
Biochem. J. 117, 641-660, 1970
A:Title: The amino acid sequences of the Fd fragments of two human gamma heavy chain
A:Reference number: A90250; MUID:70258837
A:Accession: A02091
A:Molecule type: protein
A:Residues: 1-119 <PRE>
C:Comment: This chain was obtained from IgG1 isolated from the serum of a patient wit
C:Genetics:
A:Gene: GDB:IGHV6
A:Cross-references: GDB:128528; OMIM:147070
A:Map position: 14q32.33-14q32.33
C:Superfamily: immunoglobulin V region; immunoglobulin homology
C:Keywords: heterotetramer; immunoglobulin; pyroglutamic acid
F:15-99/Domain: immunoglobulin homology <IMM>
F:1/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental
F:22-97/Disulfide bonds: #status predicted
F:35-101/Disulfide bonds: #status experimental

Query Match 1.2%; Score 7; DB 1; Length 119;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 537 SGFSLSG 543
|||||
Db 25 SGFSLSG 31

RESULT 25

H83903
 hypothetical protein BH2032 [imported] - Bacillus halodurans (strain C-125)
 C:Species: Bacillus halodurans
 C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 31-Dec-2000
 C:Accession: H83903
 R:Rafami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hirai
 Nucleic Acids Res. 28, 4317-4331, 2000
 A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
 A:Reference number: AB3650; MUID:20263314
 A:Accession: H83903
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-123 <STO>
 A:Cross-references: GB:AP001514; GB:BA000004; NID:g10174613; PIDN:BA05751.1; GSPDB:GN00
 A:Experimental source: strain C-125
 C:Genetics:
 A:Gene: BH2032
 C:Superfamily: Bacillus subtilis hypothetical protein ydhg

Query Match 1.2%; Score 7; DB 2; Length 123;
 Best Local Similarity 100.0%; Pred. No. 40;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 261 ADAIAQA 267
 |||||
 Db 75 ADAIAQA 81

RESULT 26
 S73210
 ribosomal protein S12, chloroplast - red alga (Porphyra purpurea) chloroplast
 C:Species: chloroplast Porphyra purpurea
 C:Date: 19-Mar-1997 #sequence_revision 09-May-1997 #text_change 21-Jan-2000
 C:Accession: S73210
 R:Reith, M.; Munholland, J.
 Plant Mol. Biol. Rep. 13, 333-335, 1995
 A:Title: Complete nucleotide sequence of the Porphyra purpurea chloroplast genome.
 A:Reference number: S73108
 A:Accession: S73210
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-124 <REI>
 A:Cross-references: EMBL:U38804; NID:g1276652; PIDN:AAC08175.1; PID:g1276755
 A:Note: The nucleotide sequence was submitted to the EMBL Data Library, October 1995
 C:Genetics:
 A:Gene: rps12
 A:Genome: chloroplast
 C:Superfamily: Escherichia coli ribosomal protein S12
 C:Keywords: chloroplast; protein biosynthesis; ribosome
 F:89/Modified site: beta-methylthioaspartic acid (Asp) #status predicted

Query Match 1.2%; Score 7; DB 2; Length 124;
 Best Local Similarity 100.0%; Pred. No. 41;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 18 LIRGGRI 24
 |||||
 Db 81 LIRGGRI 87

RESULT 27
 C81292
 hypothetical protein Cj1460 [imported] - Campylobacter jejuni (strain NCTC 11168)
 C:Species: Campylobacter jejuni
 C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 31-Mar-2000
 C:Accession: C81292
 R:Parkhill, J.; Wren,
 Nature 403, 665-668, 2000
 A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hyp
 A:Reference number: AB1250; MUID:20150912

A:Accession: C81292
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-127 <PAR>
 A:Cross-references: GB:AL139078; GB:AL111168; NID:g96968723; PIDN:CAB73883.1; PID:g696
 A:Experimental source: serotype O2, strain NCTC 11168
 C:Genetics:
 A:Gene: Cj1460

Query Match 1.2%; Score 7; DB 2; Length 127;
 Best Local Similarity 100.0%; Pred. No. 42;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 208 EQRRLLE 214
 |||||
 Db 85 EQRRLLE 91

RESULT 28
 C75465
 conserved hypothetical protein - Deinococcus radiodurans (strain R1)
 C:Species: Deinococcus radiodurans
 C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
 C:Accession: C75465
 R:White, O.; Eisen, J.A.; Heideberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J
 S.; Smith, H.O.; Venter, J.C.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.;
 Science 286, 1571-1577, 1999
 A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
 A:Reference number: A75250; MUID:20036896
 A:Accession: C75465
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-135 <WHI>
 A:Cross-references: GB:AE001941; GB:AE000513; NID:g6458589; PIDN:AAF10458.1; PID:g645
 A:Experimental source: strain R1
 C:Genetics:
 A:Gene: DR0836
 A:Map position: 1

Query Match 1.2%; Score 7; DB 2; Length 135;
 Best Local Similarity 100.0%; Pred. No. 44;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 65 VLPGGVD 71
 |||||
 Db 101 VLPGGVD 107

RESULT 29
 C70077
 hypothetical protein yxig N17J [imported] - Bacillus subtilis
 C:Species: Bacillus subtilis
 C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 21-Jul-2000
 C:Accession: C70077; T47104
 R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber
 C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
 A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari,
 Nature 390, 249-256, 1997
 A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gal
 iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M
 Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Laplous, A.; Lardino
 A.; Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau
 y, M.; Ogawa, K.; Ogilwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portete
 Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanl
 A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Se
 akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiya
 T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasukawa, H.; Danchin, A.
 A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.
 A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis
 A:Reference number: A69580; MUID:98044033

A:Accession: C70077
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-138 <KUN>
 A:Cross-references: GB:Z99124; GB:AL009126; NID:g2636442; PIDN:CAB15955.1; PID:g2636465
 A:Experimental source: strain 168
 A:Yoshida, K.; Sano, H.; Seki, S.; Oda, M.; Fujimura, M.; Fujita, Y.
 Microbiology 141, 337-343, 1995
 A:Title: Cloning and sequencing of a 29 kb region of the *Bacillus subtilis* genome containing the *ysj* gene
 A:Reference number: Z24350; MUID:95219088
 A:Accession: T47104
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-138 <YOS>
 A:Cross-references: EMBL:D31856; NID:g603765; PIDN:BAA06659.1; PID:g603785
 A:Experimental source: strain BGS1A1
 C:Genetics:
 A:Gene: *ysj*; N17J
 C:Superfamily: *Bacillus subtilis* hypothetical protein *ysj*
 Query Match 1.2% Score 7; DB 2; Length 138;
 Best Local Similarity 100.0%; Pred. No. 45;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 27 DQSFYA 33
 Db 126 DQSFYA 132
 RESULT 30
 D69857
 conserved hypothetical protein *ykla* - *Bacillus subtilis*
 C:Species: *Bacillus subtilis*
 C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 20-Jun-2000
 C:Accession: D69857
 R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertero, C.; Bron, S.; Brouillet, S.; Bruschini, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Chao, A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
 Nature 390, 249-256, 1997
 A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gallier, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hulio, M.F.; Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois, A.; Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel, Y.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle, R.; Rivolet, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon, A.; Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seron, A.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.
 A:Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
 A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.
 A:Reference number: A69580; MUID:98044033
 A:Accession: D69857
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-141 <KUN>
 A:Cross-references: GB:Z99110; GB:AL009126; NID:g2633472; PIDN:CAB13171.1; PID:g2633668
 A:Experimental source: strain 168
 C:Genetics:
 A:Gene: *ykla*
 C:Superfamily: hypothetical protein *ykla*
 Query Match 1.2% Score 7; DB 2; Length 141;
 Best Local Similarity 100.0%; Pred. No. 46;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 151 ELEALVK 157
 Db 111 ELEALVK 117
 RESULT 31

S53010
 RCC2 protein - rice
 C:Species: *Oryza sativa* (rice)
 C:Date: 14-Jul-1995 #sequence_revision 03-Nov-1995 #text_change 18-Feb-2000
 A:Accession: S53010
 R:Xu, Y.; Buchholz, W.G.; DeRose, R.T.; Hall, T.C.
 Plant Mol. Biol. 27, 237-248, 1995
 A:Title: Characterization of a rice gene family encoding root-specific proteins.
 A:Reference number: S53010; MUID:95195153
 A:Accession: S53010
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-146 <XUY>
 C:Superfamily: hydroxyproline-rich glycoprotein
 Query Match 1.2% Score 7; DB 2; Length 146;
 Best Local Similarity 100.0%; Pred. No. 47;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 441 GAPAVVI 447
 Db 32 GAPAVVI 38
 RESULT 32
 S53011
 RCg2 protein - rice
 C:Species: *Oryza sativa* (rice)
 C:Date: 14-Jul-1995 #sequence_revision 03-Nov-1995 #text_change 21-Jul-2000
 A:Accession: S53011
 R:Xu, Y.; Buchholz, W.G.; DeRose, R.T.; Hall, T.C.
 Plant Mol. Biol. 27, 237-248, 1995
 A:Title: Characterization of a rice gene family encoding root-specific proteins.
 A:Reference number: S53010; MUID:95195153
 A:Accession: S53011
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-146 <XUY>
 A:Cross-references: GB:L27209; NID:g786129; PIDN:AAA65512.1; PID:g786130
 C:Superfamily: hydroxyproline-rich glycoprotein
 Query Match 1.2% Score 7; DB 2; Length 146;
 Best Local Similarity 100.0%; Pred. No. 47;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 441 GAPAVVI 447
 Db 32 GAPAVVI 38
 RESULT 33
 T16440
 hypothetical protein F53A9.9 - *Caenorhabditis elegans*
 C:Species: *Caenorhabditis elegans*
 C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 24-Nov-1999
 A:Accession: T16440
 R:Miller, N.
 submitted to the EMBL Data Library, March 1995
 A:Description: The sequence of C. elegans cosmid F53A9.
 A:Reference number: Z18513
 A:Accession: T16440
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-147 <MIL>
 A:Cross-references: EMBL:U23523; NID:g746551; PID:g746560; PIDN:AAC46564.1; CESP:F53A
 A:Experimental source: strain Bristol N2
 C:Genetics:
 A:Gene: CESP:F53A9.9
 C:Superfamily: homeotic protein sal

Query Match 1.2% Score 7; DB 2; Length 147;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 67 PGGVVDVH 73
|||||||
Db 110 PGGVVDVH 116

RESULT 34

T30616
hypothetical protein 14R - Molluscum contagiosum virus 1
N:Alternate names: MC014R
C:Species: Molluscum contagiosum virus 1
C:Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 16-Feb-2001
C:Accession: T30616
R:Senkevich, T.G.; Bugert, J.J.; Sisler, J.R.; Koonin, E.V.; Darai, G.; Moss, B.
Science 273, 813-816, 1996
A:Title: Genome sequence of a human tumorigenic poxvirus: Prediction of specific host re
A:Reference number: 220876; MUID:96325459
A:Accession: T30616
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-147 <SEN>
A:Cross-references: EMBL:U60315; NID:g1491943; PIDN:AAC55142.1; PID:g1491957
C:Genetics:
A:Note: MC014R
C:Superfamily: Molluscum contagiosum virus 1 hypothetical protein 14R

Query Match 1.2% Score 7; DB 2; Length 147;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 560 APPGGRS 566
|||||||
Db 120 APPGGRS 126

RESULT 35

S77594
type 4 fimbrial biogenesis protein PilV PA4551 [imported] - Pseudomonas aeruginosa (stra
C:Species: Pseudomonas aeruginosa
C:Date: 24-Oct-1998 #sequence_revision 24-Oct-1998 #text_change 06-Oct-2000
C:Accession: S77594; A83076
R:Alm, R.A.; Mattick, J.S.
Mol. Microbiol. 16, 485-496, 1995
A:Title: Identification of a gene, pilV, required for type 4 fimbrial biogenesis in Pseu
A:Reference number: S77594; MUID:96037793
A:Accession: S77594
A:Molecule type: DNA
A:Residues: 1-185 <ALM>
A:Cross-references: EMBL:L36117
A:Experimental source: strain PA01
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; B
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim
; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A:Reference number: A82950; MUID:20437337
A:Accession: A83076
A:Molecule type: DNA
A>Status: preliminary
A:Residues: 1-185 <STO>
A:Cross-references: GB:AE004869; GB:AE004091; NID:g9950793; PIDN:AAG07939.1; GSPDB:GN001
A:Experimental source: strain PA01
C:Genetics:
A:Gene: pilV; PA4551
C:Function:
A:Description: involved in biogenesis of type 4 fimbriae
C:Keywords: fimbria; membrane protein

Query Match 1.2% Score 7; DB 2; Length 185;
Best Local Similarity 100.0%; Pred. No. 58;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 534 LHQSGFS 540
|||||||
Db 10 LHQSGFS 16

RESULT 36

S76398
hypothetical protein - Synecocystis sp. (strain PCC 6803)
C:Species: Synecocystis sp.
A:Variety: PCC 6803
C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 08-Oct-1999
C:Accession: S76398
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima,
O, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yas
DNA Res. 3, 109-136, 1996
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synecocys
S:Reference number: S74322; MUID:97061201
A:Accession: S76398
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-191 <KAN>
A:Cross-references: EMBL:D90915; GB:AB001339; NID:g1653604; PIDN:BAAL8527.1; PID:d101
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

Query Match 1.2% Score 7; DB 2; Length 191;
Best Local Similarity 100.0%; Pred. No. 60;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 12 ITSDDL 18
|||||||
Db 156 ITSDDL 162

RESULT 37

F81934
probable lipoprotein NMA0883 [imported] - Neisseria meningitidis (strain Z2491 serogr
C:Species: Neisseria meningitidis
C:Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001
C:Accession: F81934
R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Mo
; Holroyd, S.; Jagers, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandre
Nature 404, 502-506, 2000
A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491
A:Reference number: A81775; MUID:20222556
A:Accession: F81934
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-197 <PAR>
A:Cross-references: GB:AL162754; GB:AL157959; NID:g7379424; PIDN:CAB84163.1; PID:g737
A:Experimental source: serogroup A, strain Z2491
C:Genetics:
A:Superfamily: Neisseria meningitidis probable lipoprotein NMA0883

Query Match 1.2% Score 7; DB 2; Length 197;
Best Local Similarity 100.0%; Pred. No. 62;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 125 ERADSAA 131
|||||||
Db 110 ERADSAA 116

RESULT 38

T33826
hypothetical protein F23C8.6 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T33826
R:Miller, N.; Wamsley, P.
submitted to the EMBL Data Library, November 1998
A:Description: The sequence of C. elegans cosmid F23C8.
A:Reference number: 221419
A:Accession: T33826
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-205 <MWL>
A:Cross-references: EMBL:AF106583; PIDN:AD03134.1; GSPDB:GN00019; CESP:F23C8.6
A:Experimental source: strain Bristol N2; clone F23C8
C:Genetics:
A:Gene: CESP:F23C8.6
A:Map position: 1
A:Introns: 12/1; 47/3

Query Match 1.2%; Score 7; DB 2; Length 205;
Best Local Similarity 100.0%; Pred. No. 64;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 195 AOVHAEN 201
|||||
DB 54 AOVHAEN 60

RESULT 39
A38594
troponin I - fruit fly (Drosophila melanogaster) (clone E2)
C:Species: Drosophila melanogaster
C:Date: 16-Sep-1992 #sequence_revision 16-Sep-1992 #text_change 13-Aug-1999
C:Accession: A38594
R:Barbas, J.A.; Galceran, J.; Krah-Jentgens, I.; de la Pompa, J.L.; Canal, I.; Pongs, O.
Genes Dev. 5, 132-140, 1991
A:Title: Troponin I is encoded in the haplolethal region of the Shaker gene complex of D.
A:Reference number: A38594; MUID:91115093
A:Accession: A38594
A:Molecule type: DNA
A:Residues: 1-208 <BAR>
A:Cross-references: GB:X58188; NID:g2511643; PIDN:CAA41171.1; PID:g8738
C:Genetics:
A:Gene: FlyBase:wupA
A:Cross-references: FlyBase:FBgn0004028
A:Introns: 5/3; 123/3; 185/3
C:Superfamily: troponin I

Query Match 1.2%; Score 7; DB 2; Length 208;
Best Local Similarity 100.0%; Pred. No. 65;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 229 EEVEAEA 235
|||||
DB 202 EEVEAEA 208

RESULT 40
A40547
troponin I - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C:Date: 28-May-1992 #sequence_revision 28-May-1992 #text_change 13-Aug-1999
C:Accession: A40547
R:Beall, C.J.; Fyrberg, E.
J. Cell Biol. 114, 941-951, 1991
A:Title: Muscle abnormalities in Drosophila melanogaster heldup mutants are caused by mi
A:Reference number: A40547; MUID:91340840
A:Accession: A40547
A:Status: preliminary; nucleic acid sequence not shown; not compared with conceptual tra
A:Molecule type: mRNA
A:Residues: 1-208 <BEA>
A:Cross-references: GB:X59376; NID:g8057; PIDN:CAA42020.1; PID:g8058

C:Genetics:
A:Gene: FlyBase:wupA
A:Cross-references: FlyBase:FBgn0004028
C:Superfamily: troponin I

Query Match 1.2%; Score 7; DB 2; Length 208;
Best Local Similarity 100.0%; Pred. No. 65;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 229 EEVEAEA 235
|||||
DB 202 EEVEAEA 208

RESULT 41
E72385
hypothetical protein TM0375 - Thermotoga maritima (strain MSB8)
C:Species: Thermotoga maritima
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
C:Accession: E72385
R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hic
Garrett, M.W.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson,
C.M.
Nature 399, 323-329, 1999
A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome
A:Reference number: A72200; MUID:99287316
A:Accession: E72385
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-210 <ARN>
A:Cross-references: GB:AE001717; GB:AE000512; NID:g4980871; PIDN:RAD35462.1; PID:g498
A:Experimental source: strain MSB8
C:Genetics:
A:Gene: TM0375
C:Superfamily: Thermotoga maritima hypothetical protein TM0375

Query Match 1.2%; Score 7; DB 2; Length 210;
Best Local Similarity 100.0%; Pred. No. 65;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 155 LVKEGV 161
|||||
DB 128 LVKEGV 134

RESULT 42
H84244
p-nitrophenyl phosphatase [imported] - Halobacterium sp. NRC-1
C:Species: Halobacterium sp. NRC-1
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: H84244
R:Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky
; Leithauser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Ja
Jung, K.H.; Alam, M.; Freitas, T.
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.;
A:Title: Genome sequence of Halobacterium species NRC-1.
A:Reference number: A84160; MUID:20504483
A:Accession: H84244
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-212 <STO>
A:Cross-references: GB:AE004437; NID:gl0580446; PIDN:AAG19324.1; GSPDB:GN00138
C:Genetics:
A:Gene: pho2

Query Match 1.2%; Score 7; DB 2; Length 212;
Best Local Similarity 100.0%; Pred. No. 66;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 510 AKPGSGA 516
| | | | |
DB 114 AKPGSGA 120

RESULT 43

S35564
sex-determining protein SRY - common marmoset
C:Species: Callithrix jacchus (common marmoset)
C:Date: 20-May-1994 #sequence_revision 26-May-1995 #text_change 08-Sep-1997
C:Accession: S35564
R:Whitfield, L.S.; Lovell-Badge, R.; Goodfellow, P.N.
Nature 364, 713-715, 1993
A:Title: Rapid sequence evolution of the mammalian sex-determining gene SRY.
A:Reference number: S35558; MUID:93361117
A:Accession: S35564
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-227 <WHI>
C:Genetics:
A:Gene: SRY
A:Map position: Y
C:Superfamily: unassigned HMG box proteins; HMG box homology
C:Keywords: DNA binding
F:57-132/Domain: HMG box homology <HMG>

Query Match 1.2%; Score 7; DB 2; Length 227;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 304 SPPVNP 310
| | | | |
DB 187 SPPVNP 193

RESULT 44

T35123
SOS response regulator lexA - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C:Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 11-Jan-2000
C:Accession: T35123
R:Seeger, K.J.; Harris, D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, March 1998
A:Reference number: 221568
A:Accession: T35123
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-234 <SEE>
A:Cross-references: EMBL:AL022268; PIDN:CAA18339.1; GSPDB:GN00070; SCOEDB:SC4H2.24C
A:Experimental source: strain A3(2)
C:Genetics:
A:Gene: lexA; SCOEDB:SC4H2.24C
C:Superfamily: lexA protein

Query Match 1.2%; Score 7; DB 2; Length 234;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 199 AENGDIV 205
| | | | |
DB 178 AENGDIV 184

RESULT 45

F70405
hypothetical protein aq.1223 - Aquifex aeolicus
C:Species: Aquifex aeolicus
C:Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 05-Nov-1999
C:Accession: F70405
R:Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Ov

Nature 392, 353-358, 1998
A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A:Reference number: A70300; MUID:98196666
A:Accession: F70405
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-235 <AQF>
A:Cross-references: GB:AE000729; NID:g2983659; PIDN:AAC07236.1; PID:g2983670; GB:AE00
A:Experimental source: strain VF5
C:Genetics:
A:Gene: aq_1223

Query Match 1.2%; Score 7; DB 2; Length 235;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 149 KEELEAL 155
| | | | |
DB 37 KEELEAL 43

Search completed: July 30, 2001, 11:43:47
Job time: 141 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2001, 11:42:26 ; Search time 11.32 seconds
(without alignments)
1730.931 Million cell updates/sec

Title: US-09-367-496-8

Perfect score: 572

Sequence: 1 MSFQGGKSPRITSDRLIR.....RTAQKIMAPGGGRSNTLS 572

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 93435 seqs, 34255486 residues

Word size : 0

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	450	78.7	572	1	DPY4_HUMAN
2	55	9.6	564	1	DPY4_RAT
3	50	8.7	572	1	DPY4_MOUSE
4	31	5.4	572	1	DPY2_BOVIN
5	31	5.4	572	1	DPY2_CHICK
6	31	5.4	572	1	DPY2_HUMAN
7	31	5.4	572	1	DPY2_MOUSE
8	31	5.4	572	1	DPY2_RAT
9	18	3.1	358	1	DPY3_RAT
10	18	3.1	570	1	DPY3_HUMAN
11	18	3.1	570	1	DPY3_MOUSE
12	18	3.1	571	1	DPY3_XENLA
13	18	3.1	572	1	DPY1_HUMAN
14	18	3.1	572	1	DPY1_MOUSE
15	18	3.1	572	1	DPY1_RAT
16	14	2.4	519	1	DPY5_HUMAN
17	14	2.4	519	1	DPY5_RAT
18	11	1.9	854	1	UN33_CAEEL
19	8	1.4	309	1	YBGK_HAETN
20	8	1.4	457	1	HYDA_AGRRD
21	8	1.4	465	1	YGE2_ECOLI
22	8	1.4	520	1	DPY1_CAEEL
23	8	1.4	546	1	CHOD_STRSQ
24	8	1.4	713	1	PALY_DIGIA
25	8	1.4	837	1	ATSA_HUMAN
26	8	1.4	900	1	GLND_PSEAE
27	7	1.2	117	1	HV2B_RABIT
28	7	1.2	119	1	HV2C_HUMAN
29	7	1.2	124	1	RR12_PORPU
30	7	1.2	128	1	HSLU_SALTY
31	7	1.2	138	1	YXIG_BACSU
32	7	1.2	141	1	YKLA_BACSU
33	7	1.2	147	1	YV59_CAEEL

ALIGNMENTS

```

RESULT 1
DPY4_HUMAN
ID DPY4_HUMAN STANDARD; PRT; 572 AA.
AC Q14531; O00240;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE DIHYDROPYRIMIDINASE RELATED PROTEIN-4 (DRP-4) (ULIP4 PROTEIN).
GN DPYSL4 OR ULIP4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID:9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Hamajima N., Kato Y., Kowaki M., Wada Y., Sasasaki M., Nonaka M.;
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 1-553 FROM N.A.
RC TISSUE=Retina;
RA MEDLINE=98314496; PubMed=9652388;
RX Byk I., Ozon S., Sobel A.;
RT "The Ulip family phosphoproteins -- common and specific properties.";
RL Eur. J. Biochem. 254:14-24(1998).
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC (POTENTIAL).
CC -!- SIMILARITY: BELONGS TO THE DEHYDROPYRIMIDINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AB006713; BAA21886.1; -
CC EMBL; Y10976; CAA71872.1; -
CC HSSP; P18316; 1KRC.
CC InterPro; IPR002195; -
CC Pfam; PF00744; Dihydroorotase; 1.
CC CONFLICT 122 122 R -> Q (IN REF. 2).
SQ SEQUENCE 572 AA; 61905 MW; 3E72A33E3BED5BE9 CRC64;

```

Query Match 78.7%; Score 450; DB 1; Length 572;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 450; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 123 WRIRADSAACCDYSLHVDITRWHSIKEEALVKRGVNSFLVFMAYKDRCCSDSQMY 182
|||||
DB 123 WRIRADSAACCDYSLHVDITRWHSIKEEALVKRGVNSFLVFMAYKDRCCSDSQMY 182
|||||

QY 183 EITPSIIRDLGALAQVHAENGDIIVEEQKRLLEGITGPEGHVLSHPPEVEAEAVYRAVTI 242
|||||
DB 183 EITPSIIRDLGALAQVHAENGDIIVEEQKRLLEGITGPEGHVLSHPPEVEAEAVYRAVTI 242
|||||

068564 pseudomonas
P53633 saccharomyc
P51501 callithrix
P15951 homo sapien
P39577 bacillus su
P24144 rhizobium l
Q59112 acidaminoco
P71887 mycobacteri
P34516 caenorhabdi
P28163 strongyloce
Q26846 strongyloce
O07637 bacillus su

QY 243 AKQANCLYVTVKMSKGRADATAQAQKRGVVVFGEPITASLTGDSHYWKNWAKAAAFV 302
DB 243 AKQANCLYVTVKMSKGRADATAQAQKRGVVVFGEPITASLTGDSHYWKNWAKAAAFV 302
QY 303 TSPPPVNPDTTADHLTCLSSGDLQVTSAGHCTFTTAQKAVGKDNFALLIPECTNGIERM 362
DB 303 TSPPPVNPDTTADHLTCLSSGDLQVTSAGHCTFTTAQKAVGKDNFALLIPECTNGIERM 362
QY 363 SMWKECVASGKMDNEFVAVTSTNAKIFNFPYPRKGRVAVGSDADLVWNPKATKIISA 422
DB 363 SMWKECVASGKMDNEFVAVTSTNAKIFNFPYPRKGRVAVGSDADLVWNPKATKIISA 422
QY 423 KTHNLNVEYNIPEGVECRGAPAVVISQGRVALEDGKMFVTPGAGRFVPRKTPDFVYKRI 482
DB 423 KTHNLNVEYNIPEGVECRGAPAVVISQGRVALEDGKMFVTPGAGRFVPRKTPDFVYKRI 482
QY 483 KARNRLAEIHGVPRGLYDGPVHEVMVPAKPGSGAPARASCPGKISVPPVRNLHQSGFSL 542
DB 483 KARNRLAEIHGVPRGLYDGPVHEVMVPAKPGSGAPARASCPGKISVPPVRNLHQSGFSL 542
QY 543 GSOADHIAARTQAQIMAPPGGRSNTLS 572
DB 543 GSOADHIAARTQAQIMAPPGGRSNTLS 572

RESULT 2
ID DPY4_RAT STANDARD; PRT; 564 AA.
AC Q62951;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DE DIHYDROPYRIMIDINASE RELATED PROTEIN-4 (DRP-4) (COLLAPSPIN RESPONSE
DE MEDIATOR PROTEIN 3) (CRMP-3) (FRAGMENT).
GN DPYS14.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=96424532; PubMed=8815901;
RA Wang L., Strittmatter S.M.;
RT "A family of rat CRMP genes is differentially expressed in the
RT nervous system";
RL J. Neurosci. 16:6197-6207(1996).
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC (POTENTIAL).
CC -!- DEVELOPMENTAL STAGE: EXPRESSED TRANSIENTLY IN DEVELOPING SPINAL
CC CORD AND SELECTIVELY IN THE POSTNATAL CEREBELLUM.
CC -!- SIMILARITY: BELONGS TO THE DEHYDROPYRIMIDINASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
DB EMBL; U52103; AAB03281.1; -;
DR HSSP; P18316; IKRC.
DR InterPro; IPR002195; -;
DR Pfam; PF00744; Dihydroorotase; 1.
FT NON_TER 1
SQ SEQUENCE 564 AA; 61085 MW; 42050891CC1436D2 CRC64;

Query Match 9.6%; Score 55; DB 1; Length 564;
Best Local Similarity 100.0%; Pred. No. 4.le-49;
Matches 55; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 57 TIDAHGLMVLPGGVVHTRLQMPVLGMPADDFCOGTRKAALAGGTTMILDRHVPD 111
DB 49 TIDAHGLMVLPGGVVHTRLQMPVLGMPADDFCOGTRKAALAGGTTMILDRHVPD 103
RESULT 3
ID DPY4_MOUSE STANDARD; PRT; 572 AA.
AC O35098; O08886;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE DIHYDROPYRIMIDINASE RELATED PROTEIN-4 (DRP-4) (ULIP4 PROTEIN).
GN DPYS14 OR ULIP4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RA Hamajima N., Kato Y., Kowaki M., Wada Y., Sasaki M., Nonaka M.;
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ICR; TISSUE=Brain;
RX MEDLINE=98314496; PubMed=9652388;
RA Byk T., Ozon S., Sobel A.;
RT "The Ulip family phosphoproteins -- common and specific properties.";
RL Eur. J. Biochem. 254:14-24(1998).
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC (POTENTIAL).
CC -!- SIMILARITY: BELONGS TO THE DEHYDROPYRIMIDINASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
DB EMBL; AB006715; BAA21888.1; -;
DR EMBL; Y09079; CAA70299.1; -;
DR MGD; MGI:1349764; Dpys14.
DR InterPro; IPR002195; -;
DR Pfam; PF00744; Dihydroorotase; 1.
FT CONFLICT 125 126 ER -> DG (IN REF. 2).
FT CONFLICT 354 354 G -> V (IN REF. 2).
FT CONFLICT 420 420 F -> I (IN REF. 2).
SQ SEQUENCE 572 AA; 61961 MW; 37671129FC02C7AF CRC64;

Query Match 8.7%; Score 50; DB 1; Length 572;
Best Local Similarity 100.0%; Pred. No. 7e-44;
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 365 VWEKCVASGKMDNEFVAVTSTNAKIFNFPYPRKGRVAVGSDADLVWNP 414
DB 365 VWEKCVASGKMDNEFVAVTSTNAKIFNFPYPRKGRVAVGSDADLVWNP 414

RESULT 4
ID DPY2_BOVIN STANDARD; PRT; 572 AA.
AC O02675;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DE DIHYDROPYRIMIDINASE RELATED PROTEIN-2 (DRP-2) (NEURAL SPECIFIC PROTEIN
DE NSP60).
GN DPYS12.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

CC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 CC Bovidae; Bovinae; Bos.
 CC NCB1_TaxID=9913;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Kanata T.K.;
 RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (POTENTIAL).
 CC -1- SIMILARITY: BELONGS TO THE DEHYDROPYRIMIDINASE FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; U83278; AAB80618.1; -
 DR HSSP; P18316; IKRC.
 DR InterPro: IPR002195; -
 DR Pfam; PF00744; Dihydroorotase; 1.
 SQ SEQUENCE 572 AA; 62277 MW; 343507ACB9D91BDE CRC64;
 CC
 CC Query Match 5.4%; Score 31; DB 1; Length 572;
 CC Best Local Similarity 100.0%; Pred. No. 5e-24;
 CC Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 CC
 CC QY 276 GEPITASLGTGSHYWSKNWAKAAAFVTSPP 306
 CC ||||||||||||||||||||||||||||||||
 CC Db 276 GEPITASLGTGSHYWSKNWAKAAAFVTSPP 306
 CC
 CC RESULT 5
 CC ID DPY2_CHICK STANDARD; PRT; 572 AA.
 CC AC Q90635;
 CC DT 15-JUL-1998 (Rel. 36, Created)
 CC DT 15-JUL-1998 (Rel. 36, Last sequence update)
 CC DT 15-JUL-1998 (Rel. 36, Last annotation update)
 CC DE DIHYDROPYRIMIDINASE RELATED PROTEIN-2 (DRP-2) (COLLAPSPIN RESPONSE
 CC MEDIATOR PROTEIN CRMP-62).
 CC OS Gallus gallus (Chicken).
 CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
 CC OC Gallus.
 CC OX NCB1_TaxID=9031;
 CC RN SEQUENCE FROM N.A.
 CC RP TISSUE=Dorsal root ganglion;
 CC RX MEDLINE=95364923; PubMed=7637782;
 CC RA Goshima Y., Nakamura F., Strittmatter P., Strittmatter S.M.;
 CC RT "Collapsin-induced growth cone collapse mediated by an intracellular
 CC protein related to UNC-33.";
 CC RL Nature 376:509-514(1995).
 CC RN SEQUENCE FROM N.A.
 CC RP TISSUE=Liver;
 CC RA Zhou J., Chen Y., Gu J.R.;
 CC RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
 CC RN SEQUENCE FROM N.A.
 CC RX MEDLINE=20039612; PubMed=10574455;
 CC RA Kitamura K., Takayama M., Hamajima N., Nakanishi M., Sasaki M.,
 CC RA Endo Y., Takemoto T., Kimura H., Iwaki M., Nonaka M.;
 CC RT "Characterization of the human dihydropyrimidinase-related protein 2
 CC (DRP-2) gene.";
 CC RL DNA Res. 6:291-297(1999).
 CC CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (POTENTIAL).
 CC CC -1- TISSUE SPECIFICITY: UBIQUITOUS.
 CC CC -1- SIMILARITY: BELONGS TO THE DEHYDROPYRIMIDINASE FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; U17277; AAA93200.1; -
 DR HSSP; P18316; IKRC.
 DR InterPro: IPR002195; -
 DR Pfam; PF00744; Dihydroorotase; 1.
 SQ SEQUENCE 572 AA; 62330 MW; 85DB9E3DD5E54D8D CRC64;

Query Match 5.4%; Score 31; DB 1; Length 572;
 Best Local Similarity 100.0%; Pred. No. 5e-24;
 Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 CC
 CC QY 276 GEPITASLGTGSHYWSKNWAKAAAFVTSPP 306
 CC ||||||||||||||||||||||||||||||||
 CC Db 276 GEPITASLGTGSHYWSKNWAKAAAFVTSPP 306
 CC
 CC RESULT 6
 CC ID DPY2_HUMAN STANDARD; PRT; 572 AA.
 CC AC Q16555; O00424;
 CC DT 15-JUL-1998 (Rel. 36, Created)
 CC DT 15-JUL-1998 (Rel. 36, Last sequence update)
 CC DT 01-OCT-2000 (Rel. 40, Last annotation update)
 CC DE DIHYDROPYRIMIDINASE RELATED PROTEIN-2 (DRP-2) (COLLAPSPIN RESPONSE
 CC MEDIATOR PROTEIN 2) (CRMP-2) (N2A3).
 CC GN DPYSL2 OR CRMP2.
 CC OS Homo sapiens (Human).
 CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 CC OX NCB1_TaxID=9606;
 CC RN [1]
 CC RP SEQUENCE FROM N.A.
 CC RX MEDLINE=95364923; PubMed=7637782;
 CC RA Goshima Y., Nakamura F., Strittmatter P., Strittmatter S.M.;
 CC RT "Collapsin-induced growth cone collapse mediated by an intracellular
 CC protein related to UNC-33.";
 CC RL Nature 376:509-514(1995).
 CC RN [2]
 CC RP SEQUENCE FROM N.A.
 CC RX MEDLINE=97128821; PubMed=8973361;
 CC RA Hamajima N., Matsuda K., Sakata S., Tamaki N., Sasaki M., Nonaka M.;
 CC RT "A novel gene family defined by human dihydropyrimidinase and three
 CC related proteins with differential tissue distribution.";
 CC RL Gene 180:157-163(1996).
 CC RN [3]
 CC RP SEQUENCE FROM N.A.
 CC RX TISSUE=Liver;
 CC RA Zhou J., Chen Y., Gu J.R.;
 CC RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
 CC RN [4]
 CC RP SEQUENCE FROM N.A.
 CC RX MEDLINE=20039612; PubMed=10574455;
 CC RA Kitamura K., Takayama M., Hamajima N., Nakanishi M., Sasaki M.,
 CC RA Endo Y., Takemoto T., Kimura H., Iwaki M., Nonaka M.;
 CC RT "Characterization of the human dihydropyrimidinase-related protein 2
 CC (DRP-2) gene.";
 CC RL DNA Res. 6:291-297(1999).
 CC CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (POTENTIAL).
 CC CC -1- TISSUE SPECIFICITY: UBIQUITOUS.
 CC CC -1- SIMILARITY: BELONGS TO THE DEHYDROPYRIMIDINASE FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; U7279; AAA93202.1; -
 DR EMBL; D78013; BAA1191.1; -
 DR EMBL; U97105; AAC05793.1; -
 DR EMBL; AB020777; BAA86991.1; -
 DR EMBL; AB020764; BAA86991.1; JOINED.
 DR EMBL; AB020765; BAA86991.1; JOINED.
 DR EMBL; AB020766; BAA86991.1; JOINED.
 DR EMBL; AB020767; BAA86991.1; JOINED.

```
DR EMBL; AB020768; BAA86991.1; JOINED.
DR EMBL; AB020769; BAA86991.1; JOINED.
DR EMBL; AB020770; BAA86991.1; JOINED.
DR EMBL; AB020771; BAA86991.1; JOINED.
DR EMBL; AB020772; BAA86991.1; JOINED.
DR EMBL; AB020773; BAA86991.1; JOINED.
DR EMBL; AB020774; BAA86991.1; JOINED.
DR EMBL; AB020775; BAA86991.1; JOINED.
DR EMBL; AB020776; BAA86991.1; JOINED.
DR HSSP; P18316; 1KRC.
DR MIN; 602463; -.
DR InterPro; IPR002195; -.
DR Pfam; PF00744; Dihydroorotase; 1.
SQ SEQUENCE 572 AA; 62293 MW; 5CDB6CF7F5C308AD CRC64;

Query Match          5.4%; Score 31; DB 1; Length 572;
Best Local Similarity 100.0%; Pred. No. 5e-24;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 276 GEPITASLTGDSHYWSKNWAKAAAFVTSPP 306
   |||||||
Db 276 GEPITASLTGDSHYWSKNWAKAAAFVTSPP 306

RESULT 7
DPY2_MOUSE
ID DPY2_MOUSE STANDARD; PRT; 572 AA.
AC 008553;
DT 15-JUL-1998 (Rel. 36, Created)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE DIHYDROPYRIMIDINASE RELATED PROTEIN-2 (DRP-2) (ULIP 2 PROTEIN).
GN DYLIP2 OR ULIP2 OR CRMP2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ICR; TISSUE=Brain;
RX MEDLINE=98314496; PubMed=9652388;
RA Byk T., Ozon S., Sobel A.;
RT "The Ulip family phosphoproteins -- common and specific properties.";
RL Eur. J. Biochem. 254:14-24(1998).
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE DEHYDROPYRIMIDINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Y10339; CAA71370.1; -.
DR MGD; MGI:1349763; Dpysl2.
DR InterPro; IPR002195; -.
DR Pfam; PF00744; Dihydroorotase; 1.
SQ SEQUENCE 572 AA; 62170 MW; 7FD4E8A242ACF62D CRC64;

Query Match          5.4%; Score 31; DB 1; Length 572;
Best Local Similarity 100.0%; Pred. No. 5e-24;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 276 GEPITASLTGDSHYWSKNWAKAAAFVTSPP 306
   |||||||
Db 276 GEPITASLTGDSHYWSKNWAKAAAFVTSPP 306

RESULT 8
DPY3_RAT
ID DPY3_RAT STANDARD; PRT; 358 AA.
AC Q62952;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE DIHYDROPYRIMIDINASE RELATED PROTEIN-3 (DRP-3) (COLLAPSPIN RESPONSE
DE MEDIATOR PROTEIN 4) (CRMP-4) (FRAGMENT).
DN DPYSL3.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=96424532; PubMed=8815901;
RA Wang L., Strittmatter S.M.;
RT "A family of rat CRMP genes is differentially expressed in the
```

```
DPY2_RAT
ID DPY2_RAT STANDARD; PRT; 572 AA.
AC P47942;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE DIHYDROPYRIMIDINASE RELATED PROTEIN-2 (DRP-2) (TURNED ON AFTER
DE DIVISION, 64 KDA PROTEIN) (TOAD-64) (COLLAPSPIN RESPONSE MEDIATOR
DE PROTEIN 2) (CRMP-2).
DN DPYSL2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 136-142; 402-418; 441-450 & 499-511.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=Brain;
RX MEDLINE=96033765; PubMed=7472434;
RA Minturn J.E., Fryer H.J.L., Geschwind D.H., Hockfield S.;
RT "TOAD-64, a gene expressed early in neuronal differentiation in the
RT rat, is related to unc-33, a C. elegans gene involved in axon
RT outgrowth.";
RL J. Neurosci. 15:6757-6766(1995).
CC -1- FUNCTION: MAY HAVE A ROLE IN AXON ELABORATION.
CC -1- SUBCELLULAR LOCATION: TIGHTLY, BUT NONCOVALENTLY, ASSOCIATED WITH
CC MEMBRANES.
CC -1- DEVELOPMENTAL STAGE: EXPRESSED IMMEDIATELY AFTER NEURONAL BIRTH
CC AND IS DRAMATICALLY DOWNREGULATED IN THE ADULT.
CC -1- SIMILARITY: BELONGS TO THE DEHYDROPYRIMIDINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z46882; CAA86981.1; -.
DR InterPro; IPR002195; -.
DR Pfam; PF00744; Dihydroorotase; 1.
SQ SEQUENCE 572 AA; 62277 MW; C031F3BC038AA737 CRC64;

Query Match          5.4%; Score 31; DB 1; Length 572;
Best Local Similarity 100.0%; Pred. No. 5e-24;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 276 GEPITASLTGDSHYWSKNWAKAAAFVTSPP 306
   |||||||
Db 276 GEPITASLTGDSHYWSKNWAKAAAFVTSPP 306

RESULT 9
DPY3_RAT
ID DPY3_RAT STANDARD; PRT; 358 AA.
AC Q62952;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DE DIHYDROPYRIMIDINASE RELATED PROTEIN-3 (DRP-3) (COLLAPSPIN RESPONSE
DE MEDIATOR PROTEIN 4) (CRMP-4) (FRAGMENT).
DN DPYSL3.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=96424532; PubMed=8815901;
RA Wang L., Strittmatter S.M.;
RT "A family of rat CRMP genes is differentially expressed in the
```


DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DE DIHYDROXYRIMIDINASE RELATED PROTEIN-3 (DRP-3) (NEURAL SPECIFIC PROTEIN
 DE 1).
 GN NSPI.
 OS Xenopus laevis (African clawed frog).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
 CC Xenopodinae; Xenopus.
 OX NCBI_TaxID=8355;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RA Engel E.R., Lepperdinger G., Richter K.;
 RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: BELONGS TO THE DEHYDROXYRIMIDINASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; Y13069; CAA73509.1; -
 DR InterPro; IPR002195; -
 DR Pfam; PF00744; Dihydroorotase; 1.
 FT CONFLICT 571 AA; 62099 MW; 30A841522E27743D CRC64;
 SQ SEQUENCE 571 AA; 62099 MW; 30A841522E27743D CRC64;

 Query Match 3.1%; Score 18; DB 1; Length 571;
 Best Local Similarity 100.0%; Pred. No. 1.9e-10;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 289 HYWKNWAKAAAFVTSPP 306
 Db |||||
 Db 289 HYWKNWAKAAAFVTSPP 306

 RESULT 13
 DPYL_HUMAN
 ID DPYL_HUMAN STANDARD; PRT; 572 AA.
 AC Q14194; Q13024;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE DIHYDROXYRIMIDINASE RELATED PROTEIN-1 (DRP-1) (COLLAPSPIN RESPONSE
 DE MEDIATOR PROTEIN 1) (CRMP-1).
 GN CRMP1 OR DPYSL1.
 OS Homo sapiens (Human).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Euthera; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RA Hamajima N., Matsuda K., Sakata S., Tamaki N., Sasaki M., Nonaka M.;
 RL "A novel gene family defined by human dihydroxyrimidinase and three
 RL related proteins with differential tissue distribution.";
 RL Nature 180:157-163(1996).
 RN [2]
 RP SEQUENCE OF 64-572 FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=95364923; PubMed=7637782;
 RA Goshima Y., Nakamura F., Strittmatter P., Strittmatter S.M.;
 RT "Collapsin-induced growth cone collapse mediated by an intracellular
 RT protein related to UNC-33.";
 RL Nature 376:509-514(1995).
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (POTENTIAL).
 CC -1- TISSUE SPECIFICITY: BRAIN.

CC -1- SIMILARITY: BELONGS TO THE DEHYDROXYRIMIDINASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; D78012; BA111190.1; -
 DR EMBL; U17278; AAA93201.1; -
 DR MIN; 602462; -
 DR InterPro; IPR002195; -
 DR Pfam; PF00744; Dihydroorotase; 1.
 FT CONFLICT 504 504 Y -> H (IN REF. 2).
 SQ SEQUENCE 572 AA; 62183 MW; A5385FCC79328A30 CRC64;

 Query Match 3.1%; Score 18; DB 1; Length 572;
 Best Local Similarity 100.0%; Pred. No. 1.9e-10;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 289 HYWKNWAKAAAFVTSPP 306
 Db |||||
 Db 289 HYWKNWAKAAAFVTSPP 306

 RESULT 14
 DPYL_MOUSE
 ID DPYL_MOUSE STANDARD; PRT; 572 AA.
 AC P97427; O08554; O35097;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE DIHYDROXYRIMIDINASE RELATED PROTEIN-1 (DRP-1) (COLLAPSPIN RESPONSE
 DE MEDIATOR PROTEIN 1) (CRMP-1) (ULIP3 PROTEIN).
 GN CRMP1 OR DPYSL1 OR ULIP3.
 OS Mus musculus (Mouse).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97262103; PubMed=9107681;
 RA Cohen-Salmon M., Crozet F., Rebillard G., Petit C.;
 RT "Cloning and characterization of the mouse collapsin response
 RT mediator protein-1, Crmp1.";
 RL Mamm. Genome 8:349-351(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ICR; TISSUE=Brain;
 RX MEDLINE=98314496; PubMed=9652388;
 RA Byk T., Ozon S., Sobel A.;
 RT "The Ulip family phosphoproteins -- common and specific properties.";
 RL Eur. J. Biochem. 254:14-24(1998).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Embryo;
 RA Hamajima N., Kato Y., Kowaki M., Wada Y., Sasaki M., Nonaka M.;
 RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (POTENTIAL).
 CC -1- SIMILARITY: BELONGS TO THE DEHYDROXYRIMIDINASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U72875; AAB39703.1; -

FT VARIANT 360 360 W -> R (IN DHP DEFICIENCY).
 FT FTID-VAR_002269.
 FT VARIANT 435 435 G -> R (IN DHP DEFICIENCY).
 FT FTID-VAR_002270.
 FT VARIANT 490 490 R -> T (IN DHP DEFICIENCY).
 FT FTID-VAR_002271.
 SQ SEQUENCE 519 AA; 56629 MW; 882E33D7C49D6ECC CRC64;

Query Match 2.4%; Score 14; DB 1; Length 519;
 Best Local Similarity 100.0%; Pred. No. 2.7e-06;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 380 FVAVTSTNAKIFN 393
 DB 374 FVAVTSTNAKIFN 387
 |||||

RESULT 17

DPYS_RAT STANDARD; PRT; 519 AA.
 AC Q63150;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DE DIHYDROPYRIMIDINASE (EC 3.5.2.2) (DHPASE) (HYDANTOINASE) (DHP).
 GN DPYS.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC TISSUE=Liver;
 RX MEDLINE=96283806; PubMed=8679696;
 RA Matsuda K., Sakata S., Kaneko M., Hamajima N., Nonaka M., Sasaki M.,
 Tanaki N.;
 RT "Molecular cloning and sequencing of a cDNA encoding
 dihydropyrimidinase from the rat liver.";
 RL Biochim. Biophys. Acta 1307:140-144(1996).
 CC -!- CATALYTIC ACTIVITY: 5,6-DIHYDROURACIL + H(2)O = 3-
 UREIDOPROPIONATE.
 CC -!- SIMILARITY: BELONGS TO THE DEHYDROPYRIMIDINASE FAMILY.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; D63704; BAA09833.1; -
 DR InterPro; IPR002195; -
 DR Pfam; PF00744; Dihydroorotase; 1.
 KW Hydrolase.
 SQ SEQUENCE 519 AA; 56833 MW; CA4CEF46801FCF4B CRC64;

Query Match 2.4%; Score 14; DB 1; Length 519;
 Best Local Similarity 100.0%; Pred. No. 2.7e-06;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 91 QGTAALAGGTTMI 104
 DB 85 QGTAALAGGTTMI 98
 |||||

RESULT 18

UN33_CAEEL STANDARD; PRT; 854 AA.
 AC Q01630;
 DT 01-JUL-1993 (Rel. 26, Created)

DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE UNC-33 PROTEIN.
 GN UNC-33.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BRISTOL N2;
 RX MEDLINE=93106371; PubMed=1468626;
 RA Li W., Herman R.K., Shaw J.E.;
 RT "Analysis of the Caenorhabditis elegans axonal guidance and outgrowth
 gene unc-33.";
 RL Genetics 132:675-689(1992).
 CC -!- FUNCTION: INVOLVED IN AXONAL GUIDANCE AND OUTGROWTH.
 CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
 CC -!- ALTERNATIVE PRODUCTS: THREE PUTATIVE POLYPEPTIDES MAY ARISE BY
 CC ALTERNATIVE SPLICING.
 CC -!- DEVELOPMENTAL STAGE: UNC-33 PROTEINS ARE DISTRIBUTED EXCLUSIVELY
 CC WITHIN NEURONAL PROCESSES AFTER EARLY EMBRYOGENESIS.
 CC -!- SIMILARITY: BELONGS TO THE DEHYDROPYRIMIDINASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; Z14148; CAA78520.1; -
 DR EMBL; Z14148; CAA78521.1; -
 DR EMBL; Z14148; CAA78522.1; -
 DR EMBL; Z14146; CAA78516.1; -
 DR EMBL; Z14146; CAA78517.1; -
 DR EMBL; Z14146; CAA78518.1; -
 DR PIR; S24643; S24643.
 DR PIR; S24644; S24644.
 DR PIR; S33558; S33558.
 DR InterPro; IPR002195; -
 DR Pfam; PF00744; Dihydroorotase; 1.
 KW Alternative splicing.
 FT VARSPLIC 1 175 MISSING (IN 72.1 KDA ISOFORM).
 FT VARSPLIC 1 331 MISSING (IN 55.5 KDA ISOFORM).
 SQ SEQUENCE 854 AA; 90819 MW; A8073DDE251D2D77 CRC64;

Query Match 1.9%; Score 11; DB 1; Length 854;
 Best Local Similarity 100.0%; Pred. No. 0.0057;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 380 FVAVTSTNAK 390
 DB 683 FVAVTSTNAK 693
 |||||

RESULT 19

YBCK_HAEIN STANDARD; PRT; 309 AA.
 AC P44298;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE HYPOTHETICAL PROTEIN H11730.
 GN H11730.
 OS Haemophilus influenzae.
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
 OC Haemophilus.
 OX NCBI_TaxID=727;
 RN [1]
 RP SEQUENCE FROM N.A.

RC STRAIN=RD / KW20 / ATCC 51907;
 RX MEDLINE=95350630; PubMed=7542800;
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
 RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
 RA Scott J.D., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
 RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
 RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
 RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
 RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
 RA Venter J.C.;
 RT "Whole-genome random sequencing and assembly of Haemophilus
 RT influenzae Rd.";
 RL Science 269:496-512(1995).
 RC -1- SIMILARITY: STRONG. TO E COLI YBCK.
 CC -1- SIMILARITY: TO B.SUBTILIS YCSJ AND YEAST UREA AMIDOLYASE (DURL2).
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; U32845; AAC23376.1; -
 DR TIGR; H11730; -
 KW Hypothetical protein.
 SQ SEQUENCE 309 AA; 34562 MW; E34D87B4C838EFID CRC64;
 Query Match 1.4%; Score 8; DB 1; Length 309;
 Best Local Similarity 100.0%; Pred. No. 3.2;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 190 DLGALAAQV 197
 Db 264 DLGALAAQV 271
 RESULT 20
 HYDA_AGRD
 ID HYDA_AGRD STANDARD; PRT; 457 AA.
 AC Q44184;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DE D-HYDANTOINASE (EC 3.5.2.2) (DIHYDROPYRIMIDINASE) (DHPASE).
 OS Agrobacterium radiobacter.
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Rhizobiaceae; Agrobacterium.
 OX NCBI_TaxID=358;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NRRLB 11291;
 RA Grifantini R.;
 RL Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.
 CC -1- CATALYTIC ACTIVITY: 5,6-DIHYDROURACIL + H(2)O - 3-
 CC UREIDOPROPIONATE.
 CC -1- SIMILARITY: BELONGS TO THE DEHYDROPYRIMIDINASE FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; X91070; CAA62549.1; -
 DR InterPro: IPR002195; -
 DR Pfam: PF00744; Dihydroorotase; 1.
 KW Hydrolase.

SQ SEQUENCE 457 AA; 49939 MW; 9BC8EE0A86914A5B CRC64;
 Query Match 1.4%; Score 8; DB 1; Length 457;
 Best Local Similarity 100.0%; Pred. No. 4.5;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 401 VAVGSDAD 408
 Db 382 VAVGSDAD 389
 RESULT 21
 YGEZ_ECOLI
 ID YGEZ_ECOLI STANDARD; PRT; 465 AA.
 AC Q46806;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DE HYPOTHETICAL 51.5 KDA PROTEIN IN PBL-LYSS INTERGENIC REGION.
 GN YGEZ.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12.";
 RL Science 277:1453-1474(1997)
 CC -1- SIMILARITY: BELONGS TO THE DEHYDROPYRIMIDINASE FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; U38375; AAA83054.1; -
 DR EMBL; AB000370; AAC75911.1; -
 DR EcoGene: EGI3056; YgeZ.
 DR InterPro: IPR002195; -
 DR Pfam: PF00744; Dihydroorotase; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 465 AA; 51503 MW; F1B0C91922CE812D CRC64;
 Query Match 1.4%; Score 8; DB 1; Length 465;
 Best Local Similarity 100.0%; Pred. No. 4.5;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 67 PGQVDVHT 74
 Db 57 PGQVDVHT 64
 RESULT 22
 DPY1_CAEEL
 ID DPY1_CAEEL STANDARD; PRT; 520 AA.
 AC Q18677;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DE HYPOTHETICAL 56.3 KDA PROTEIN C47E12.8 IN CHROMOSOME IV.
 GN C47E12.8.
 OS Caenorhabditis elegans.

DR EMBL; AJ002221; CAA05251.1; -
 DR InterPro; IPR001106; -
 DR Pfam; PF00221; PAL; 1.
 DR PROSITE; PS00488; PAL_HISTIDASE; 1.
 KW Lyase; Phenylpropanoid metabolism; Multigene family.
 FT SITE 198 200 MODIFIED TO FORM 4-METHYLLIDENE-IMIDAZOLE-
 5-ONE (BY SIMILARITY).
 SQ SEQUENCE 713 AA; 77732 MW; 747CB64C65BDC6A CRC64;

Query Match 1.4%; Score 8; DB 1; Length 713;
 Best Local Similarity 100.0%; Pred. No. 6.6;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 451 RVALEDGK 458
 DB 635 RVALEDGK 642

RESULT 25
 ATSA_HUMAN STANDARD; PRT; 837 AA.
 AC 075173; Q5UN83;
 DT 01-OCT-2000 (Rel. 40, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE ADAM-TS 4 PRECURSOR (EC 3.4.24.-) (A DISINTEGRIN AND METALLOPROTEINASE
 DE WITH THROMBOSPONDIN MOTIFS 4) (ADAMTS-4) (ADAM-TS4) (AGGREGANASE 1)
 DE (ADMP-1).
 GN ADAMTS4 OR KIAA0688.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=98403880; PubMed=9734811;
 RA Ishikawa K.-I., Nagase T., Suyama M., Miyajima N., Tanaka A.,
 RA Kotani H., Nomura N., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. X.
 RT The complete sequences of 100 new cDNA clones from brain which can
 RT code for large proteins in vitro.";
 RL DNA Res. 5:169-176(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95286303; PubMed=10356395;
 RA Tortorella M.D., Burn T.C., Pratta M.A., Abbaszade I., Hollis J.M.,
 RA Liu R., Rosenfeld S.A., Copeland R.A., Decicco C.P., Wynn R.,
 RA Rockwell A., Yang F., Duke J.L., Solomon K., George H., Bruckner R.,
 RA Nagase H., Itoh Y., Ellis D.M., Ross H., Wiswall B.H., Murphy K.,
 RA Hillman M.C. Jr., Hollis G.F., Newton R.C., Magolda R.L.,
 RA Trzaskos J.M., Arner E.C.;
 RT "Purification and cloning of aggrecanase-1: a member of the ADAMTS
 RT family of proteases.";
 RL Science 284:1664-1666(1999).
 RN [3]
 RP PARTIAL SEQUENCE, AND CHARACTERIZATION.
 RX MEDLINE=20400518; PubMed=10827174;
 RA Tortorella M., Pratta M., Liu R.O., Abbaszade I., Ross H., Burn T.,
 RA Arner E.;
 RT "The thrombospondin motif of aggrecanase-1 (ADAMTS-4) is critical for
 RT aggrecan substrate recognition and cleavage.";
 RJ J. Biol. Chem. 275:25791-25797(2000).
 CC -1- FUNCTION: CLEAVES AGGREGAN, A CARTILAGE PROTEOGLYCAN, AND MAY BE
 CC INVOLVED IN ITS TURNOVER. MAY PLAY AN IMPORTANT ROLE IN THE
 CC DESTRUCTION OF AGGREGAN IN ARTHRITIC DISEASES.
 CC -1- CATALYTIC ACTIVITY: CLEAVES AGGREGAN AT THE 392-GLU-|-ALA-393
 CC SITE.
 CC -1- COPAFCTOR: BINDS ONE ZINC ION (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: SECRETED. ASSOCIATED WITH THE EXTRACELLULAR
 CC MATRIX (BY SIMILARITY).
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN BRAIN, LUNG AND HEART. EXPRESSED

CC AT VERY LOW LEVEL IN PLACENTA AND SKELETAL MUSCLES.
 CC -1- INDUCTION: BY INTERLEUKIN 1.
 CC -1- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE 1 DOMAINS ARE IMPORTANT
 CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
 CC -1- PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B (ZINC
 CC METALLOPROTEASE); ALSO KNOWN AS THE REPROLYSIN SUBFAMILY.
 CC -1- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 TSP TYPE-1 DOMAIN.
 CC -1- CAUTION: HAS SOMETIMES BEEN REFERRED TO ADAMTS2.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AE014588; BAA31663.1; -
 CC EMBL; AF148213; AAD41494.1; -
 CC MIM; 603876; -
 CC HSSP; P34179; IIAG.
 DR MEROPS; M12.221; -
 DR InterPro; IPR000130; -
 DR InterPro; IPR000884; -
 DR InterPro; IPR001590; -
 DR Pfam; PF00090; tsp_1; 1.
 DR Pfam; PF01421; Reprolysin; 1.
 DR PROSITE; PS00142; ZINC_PROTEASE; 1.
 DR PROSITE; PS0215; ADAM_MEPRO; 1.
 DR PROSITE; PS00092; TSP1; 1.
 DR PROSITE; PS00427; DISINTEGRINS; FALSE_NEG.
 KW Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
 KW Extracellular matrix.
 FT SIGNAL 1 51 POTENTIAL.
 FT PROPEP 52 212
 FT CHAIN 213 837 ADAM-TS 4.
 FT SITE 194 194 CYSTEINE SWITCH (POTENTIAL).
 FT METAL 361 361 ZINC (CATALYTIC) (BY SIMILARITY).
 FT ACT_SITE 362 362 BY SIMILARITY.
 FT METAL 365 365 ZINC (CATALYTIC) (BY SIMILARITY).
 FT METAL 371 371 ZINC (CATALYTIC) (BY SIMILARITY).
 FT DOMAIN 437 519 DISINTEGRIN-LIKE.
 FT DOMAIN 520 576 TSP-TYPE 1 1.
 FT DOMAIN 577 685 CYS-RICH.
 FT DOMAIN 686 837 SPACER.
 FT DOMAIN 247 252 POLY-ALA.
 FT CARBOHYD 68 68 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CONFLICT 77 77 A -> T (IN REF. 1).
 SQ SEQUENCE 837 AA; 90224 MW; 5DF9C9AC137DF41F CRC64;

Query Match 1.4%; Score 8; DB 1; Length 837;
 Best Local Similarity 100.0%; Pred. No. 7.6;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 512 PGSGAPAR 519
 DB 73 PGSGAPAR 80

RESULT 26
 GLND_PSEAE STANDARD; PRT; 900 AA.
 ID GLND_PSEAE
 AC Q929HO;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 01-OCT-2000 (Rel. 40, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE [PROTEIN-PII] URIDYLTRANSFERASE (EC 2.7.7.59) (PII URIDYL-
 DE TRANSFERASE) (URIDYL REMOVING ENZYME) (UTASE).
 GN GLND OR PA3658.
 OS Pseudomonas aeruginosa.

CC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
CC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-PA01;
RA Nashimoto H.;
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
Garber R.L., Goltzy R., Tolentino E., Westbrook-Wadman S., Yuan Y.,
Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
RA "Complete genome sequence of Pseudomonas aeruginosa PA01, an
FT opportunistic pathogen.";
RL Nature 406:959-964(2000).
CC -1- FUNCTION: MODIFIES, BY URIDYLATION OR DEURIDYLATION THE PII
CC (GLNB) REGULATORY PROTEIN (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: UTP + [PROTEIN-PII] = PYROPHOSPHATE +
CC URIDYL-[PROTEIN-PII].
CC -1- SIMILARITY: BELONGS TO THE GLND FAMILY.
CC -1- CAUTION: REF.1 DIFFERS FROM THAT SHOWN DUE TO A FRAMESHIFT IN
CC POSITION 285.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB024601; BAA75913.1; ALT_FRAME.
DR EMBL; AE004785; AAG07046.1; -
DR InterPro; IPR002912; -
DR InterPro; IPR002934; -
DR Pfam; PF01842; ACT; 1.
DR Pfam; PF01909; NTP_transf_2; 1.
DR Pfam; PF01909; Nucleotidyltransferase.
KW Transferase; Nucleotidyltransferase.
FT CONFLICT 811 811 L -> V (IN REF. 1).
SQ SEQUENCE 900 AA; 103404 MW; 7C71F31EC284E836 CRC64;

Query Match 1.4%; Score 8; DB 1; Length 900;
Best Local Similarity 100.0%; Pred. No. 8.2;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 326 LQVTSAAH 333
Db 171 LQVTSAAH 178
|||||||

RESULT 27
HV2B_RABIT
ID HV2B_RABIT STANDARD; PRT; 117 AA.
AC P01828;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE IG HEAVY CHAIN V-A2 REGION K-25.
OS Oryctolagus cuniculus (Rabbit).
CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE.
RX MEDLINE=76039436; PubMed=241319;
RA Jaton J.-C.;

RT "Comparison of the amino acid sequences of the variable domains of
RT two homogeneous rabbit antibodies to type III pneumococcal
RL polysaccharide.";
RL Biochem. J. 147:235-247(1975).
CC -1- MISCELLANEOUS: THIS CHAIN WAS OBTAINED FROM ANTIBODY TO TYPE III
CC PNEUMOCOCCI AND WAS ISOLATED FROM THE SERUM OF A SINGLE RABBIT.
DR PIR; A02104; GARB2K.
DR InterPro; IPR003006; -
DR Pfam; PF00047; ig; 1.
KW Immunoglobulin V region.
FT MOD_RES 1 1 PYRROLIDONE CARBOXYLIC ACID.
FT DISULFID 21 91
FT NON_TER 117 117
SQ SEQUENCE 117 AA; 12580 MW; 28DD87FDB7AEE9B8 CRC64;

Query Match 1.2%; Score 7; DB 1; Length 117;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 537 SGFSLSG 543
Db 24 SGFSLSG 30
|||||||

RESULT 28
HV2C_HUMAN
ID HV2C_HUMAN STANDARD; PRT; 119 AA.
AC P01816;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE IG HEAVY CHAIN V-II REGION DAW.
OS Homo sapiens (Human).
CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE.
RX MEDLINE=70258837; PubMed=5449120;
RA Press E.M., Hogg N.M.;
RA "The amino acid sequences of the Fd fragments of two human gamma-1
RT heavy chains.";
RL Biochem. J. 117:641-660(1970).
CC -1- MISCELLANEOUS: THIS CHAIN WAS OBTAINED FROM IGG1 ISOLATED FROM THE
CC SERUM OF A PATIENT WITH HYPERGAMMAGLOBULINEMIA.
DR PIR; A02091; GIHUDW.
DR InterPro; IPR003006; -
DR Pfam; PF00047; ig; 1.
KW Immunoglobulin V region.
FT MOD_RES 1 1 PYRROLIDONE CARBOXYLIC ACID.
FT NON_TER 119 119
SQ SEQUENCE 119 AA; 13045 MW; 4E13E00214BAD789 CRC64;

Query Match 1.2%; Score 7; DB 1; Length 119;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 537 SGFSLSG 543
Db 25 SGFSLSG 31
|||||||

RESULT 29
RR12_PORPU
ID RR12_PORPU STANDARD; PRT; 124 AA.
AC P51289;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE CHLOROPLAST 30S RIBOSOMAL PROTEIN S12.
GN RPS12.

OS Porphyra purpurea.
 OG Chloroplast.
 CC Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Bangiaceae; Porphyra.
 CC NCBI_TaxID=2787;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AVONPORT;
 RA Reith M.E., Munnolland J.;
 RT "Complete nucleotide sequence of the Porphyra purpurea chloroplast
 genome.";
 RL Plant Mol. Biol. Rep. 13:333-335(1995).
 CC -1- SIMILARITY: BELONGS TO THE S12P FAMILY OF RIBOSOMAL PROTEINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U38804; AAC08175.1; -;
 DR Mendel; 10344; PORPU; rps12; 1.
 DR InterPro; IPR000230; -;
 DR Pfam; PF00164; Ribosomal_S12; 1.
 DR PRINTS; PRO1034; RIBOSOMALS12.
 DR PROSITE; PS00055; RIBOSOMAL_S12; 1.
 KW Ribosomal protein; Chloroplast.
 SQ SEQUENCE 124 AA; 13904 MW; 6632BE623CA11FB1 CRC64;

Query Match 1.2%; Score 7; DB 1; Length 124;
 Best Local Similarity 100.0%; Pred. No. 16;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 LIRGGRI 24
 |||||
 DB 81 LIRGGRI 87

RESULT 30

HSLU_SALTY
 ID HSLU_SALTY STANDARD; PRT; 128 AA.
 AC O30911;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE ATP-DEPENDENT HSL PROTEASE ATP-BINDING SUBUNIT HSLU (HEAT SHOCK
 DE PROTEIN HSLU) (FRAGMENT).
 GN HSLU.
 OS Salmonella typhimurium.
 CC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 CC Salmonella.
 CC NCBI_TaxID=602;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97451029; PubMed=9302299;
 RA Valdivia R.H., Falkow S.;
 RT "Fluorescence-based isolation of bacterial genes expressed within
 RT host cells.";
 RL Science 277:2007-2011(1997).
 CC -1- FUNCTION: CHAPERONE SUBUNIT OF A PROTEASOME-LIKE DEGRADATION
 CC COMPLEX (BY SIMILARITY).
 CC -1- SUBUNIT: INTERACTS WITH HSLV (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE CLXP CHAPERONE FAMILY. HSLU SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

CC or send an email to license@isb-sib.ch.
 CC -----

DR EMBL; AF020812; AAB80745.1; -;
 DR StyGene; SG10655; hslu.
 KW Chaperone; ATP-binding; Heat shock.
 FT NON_TER 1
 FT NON_TER 128
 SQ SEQUENCE 128 AA; 14066 MW; 2378E66D9113BE41 CRC64;

Query Match 1.2%; Score 7; DB 1; Length 128;
 Best Local Similarity 100.0%; Pred. No. 16;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 557 KIMAPPG 563
 |||||
 DB 13 KIMAPPG 19

RESULT 31

YXIG_BACSU
 ID YXIG_BACSU STANDARD; PRT; 138 AA.
 AC P42299;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 01-NOV-1995 (Rel. 32, Last annotation update)
 DE HYPOTHETICAL 16.2 KDA PROTEIN IN WAPA-LICT INTERGENIC REGION.
 GN YXIG OR N17J.
 OS Bacillus subtilis.
 CC Bacteria; Firmicutes; Bacillus/Clostridium group;
 CC Bacillus/Staphylococcus group; Bacillus.
 CC NCBI_TaxID=1423;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168 / BGSC1A1;
 RX MEDLINE=95219088; PubMed=7704263;
 RA Yoshida K.-I., Sano H., Seki S., Oda M., Fujimura M., Fujita Y.;
 RT "Cloning and sequencing of a 29 kb region of the Bacillus subtilis
 RT genome containing the hut and wapa loci.";
 RL Microbiology 141:337-343(1995).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; D11856; BAA06659.1; -;
 DR EMBL; D83026; BAA11686.1; -;
 DR EMBL; D29885; BAA06263.1; -;
 DR EMBL; Z59124; CAB15955.1; -;
 DR Subtilist; BG11136; yxig.
 KW Hypothetical protein.
 SQ SEQUENCE 138 AA; 16228 MW; DB806FB61EAF4D94 CRC64;

Query Match 1.2%; Score 7; DB 1; Length 138;
 Best Local Similarity 100.0%; Pred. No. 17;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 27 DQCSFYA 33
 |||||
 DB 126 DQCSFYA 132

RESULT 32

YKLA_BACSU
 ID YKLA_BACSU STANDARD; PRT; 141 AA.
 AC O34762;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)

DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE HYPOTHETICAL 14.9 KDA PROTEIN IN PROA-METC INTERGENIC REGION.
GN YKLA.

OS Bacillus subtilis.

CC Bacteria; Firmicutes; Bacillus/clostridium group;

CC Bacillus/Staphylococcus group; Bacillus.

OX NCBI_TaxID=1423;

RN (1)

RP SEQUENCE FROM N.A.

RC STRAIN-168;

RA Devine K.M.;

RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.

CC -1- SIMILARITY: BELONGS TO THE OSMC/OHR FAMILY.

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; AJ002571; CAA05593.1; -

DR EMBL; Z99110; CAB13171.1; -

DR Subtilist; BG13238; YKLA.

KW Hypothetical protein.

SQ SEQUENCE 141 AA; 14868 MW; 7F4A32D0619D2233 CRC64;

Query Match 1.2%; Score 7; DB 1; Length 141;

Best Local Similarity 100.0%; Pred. No. 17;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 151 ELEALVK 157

DB 111 ELEALVK 117

RESULT 33

ID YV59 CAEEL

AC P50439; STANDARD; PRT; 147 AA.

DT 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 01-OCT-1996 (Rel. 34, Last annotation update)

DE HYPOTHETICAL 15.2 KDA PROTEIN F53A9.9 IN CHROMOSOME X.

GN F53A9.9

OS Caenorhabditis elegans.

CC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;

CC Rhabditidae; Peloderinae; Caenorhabditis.

OX NCBI_TaxID=6239;

RN (1)

RP SEQUENCE FROM N.A.

RC STRAIN-BRISTOL N2;

RA Miller N.;

RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; U23523; AAC46564.1; -

DR Wormpep; F53A9.9; CE02769.

KW Hypothetical protein.

SQ SEQUENCE 147 AA; 15184 MW; E08CEE80E91ADFCD CRC64;

Query Match

Best Local Similarity 100.0%; Pred. No. 18;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 67 PGVDVH 73

DB 110 PGVDVH 116

RESULT 34

ID OMLA_PSEFL

AC O68564; STANDARD; PRT; 175 AA.

DT 30-MAY-2000 (Rel. 39, Created)

DT 30-MAY-2000 (Rel. 39, Last sequence update)

DT 30-MAY-2000 (Rel. 39, Last annotation update)

DE OUTER MEMBRANE LIPOPROTEIN OMLA PRECURSOR.

GN OMLA OR OPRX.

OS Pseudomonas fluorescens.

CC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;

CC Pseudomonas.

OX NCBI_TaxID=294;

RN (1)

RP SEQUENCE FROM N.A.

RC STRAIN-ATCC 15453;

RX MEDLINE=99138728; PubMed=9973334;

RA Ochsner U.A.; Vasil A.I.; Johnson Z.; Vasil M.L.;

RT "Pseudomonas aeruginosa fur overlaps with a gene encoding a novel

RT outer membrane lipoprotein, OmlA.";

RL J. Bacteriol. 181:1099-1109(1999).

CC -1- FUNCTION: MAY HAVE A STRUCTURAL ROLE IN MAINTAINING THE CELL

CC ENVELOPE INTEGRITY.

CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE OUTER MEMBRANE BY A LIPID

CC ANCHOR (BY SIMILARITY).

CC -1- SIMILARITY: BELONGS TO THE SMPA/OMLA FAMILY.

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; AF050677; AAC05681.1; -

KW Outer membrane; Lipoprotein; Signal.

FT SIGNAL 1 21 BY SIMILARITY.

FT CHAIN 22 175 OUTER MEMBRANE LIPOPROTEIN OMLA.

FT LIPID 22 22 N-ACYL DIGLYCERIDE (PROBABLE).

FT DOMAIN 163 175 PRO-RICH.

SQ SEQUENCE 175 AA; 19123 MW; B03950C6931F57E7 CRC64;

Query Match 1.2%; Score 7; DB 1; Length 175;

Best Local Similarity 100.0%; Pred. No. 21;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 508 VPAKPGS 514

DB 141 VPAKPGS 147

RESULT 35

ID YIPC_YEAST

AC F53633; STANDARD; PRT; 176 AA.

DT 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 15-DEC-1998 (Rel. 37, Last annotation update)

DE YIP3 PROTEIN.

GN YIP3 OR YNL044W OR N2650.

OS Saccharomyces cerevisiae (Baker's yeast).

CC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

CC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.

OX NCBI_TaxID=4932;


```
RN RP SEQUENCE FROM N.A.
RC STRAIN-S288C / FY1679;
RX MEDLINE=96310629; PubMed=8740423;
RT "The sequence of 12.8 kb from the left arm of chromosome XIV reveals
RT a sigma element, a pro-tRNA and six complete open reading frames, one
RT of which encodes a protein similar to the human leukotriene A4
RT hydrolase.";
RL Yeast 12:493-499(1996).
RN [2]
RP SEQUENCE FROM N.A.
RA Matern H.T., Gallwitz D.;
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Duesterhoeft A., Floeth M., Fritz C., Heuss-Neitzel D., Hilbert H.,
RL Moestl D.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X97401; CAA66058.1; -
DR EMBL; X97400; CAA66057.1; -
DR EMBL; X94547; CAA64238.1; ALT_SEQ.
DR EMBL; Z71320; CAA95911.1; ALT_SEQ.
DR EMBL; Z71321; CAA95913.1; -
DR SGD; S0004989; YIP3.
KW Transmembrane.
FT TRANSMEM 84 104 POTENTIAL.
FT TRANSMEM 129 149 POTENTIAL.
FT SEQUENCE 176 AA; 19446 MW; C88EA166EEDDE11B8 CRC64;
SQ
Query Match 1.2%; Score 7; DB 1; Length 176;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 191 LGALAQV 197
Db 4 LGALAQV 10
RESULT 36
SRY_CALJA
ID SRY_CALJA STANDARD; PRT; 227 AA.
AC P51501;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE SEX-DETERMINING REGION Y PROTEIN (TESTIS-DETERMINING FACTOR).
GN SRY.
OS Callithrix jacchus (Common marmoset).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Platyrrhini; Callitrichidae;
OC Callithrix.
OX NCBI_TaxID=9483;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93361117; PubMed=8355783;
RA Whitfield L.S., Lovell-Badge R., Goodfellow P.N.;
RL "Rapid sequence evolution of the mammalian sex-determining gene SRY.";
RT Nature 364:713-715(1993).
CC -1- FUNCTION: TRANSCRIPTIONAL ACTIVATOR WHICH REGULATES A GENETIC
CC SWITCH IN MALE DEVELOPMENT. IT IS RESPONSIBLE FOR INITIATING MALE
CC SEX DETERMINATION. SRY HMG BOX RECOGNIZES DNA BY PARTIAL
```

```
CC INTERCALATION IN THE MINOR GROOVE.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- SIMILARITY: CONTAINS 1 HMG BOX.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X86386; CAA60146.1; -
DR HSSP; Q05066; LHRZ.
DR InterPro; IPR000910; -
DR Pfam; PF00505; HMG_box; 1.
KW DNA-binding; Nuclear protein; Transcription regulation; Activator;
KW Sexual differentiation.
FT DNA_BIND 60 128 HMG_BOX.
FT SEQUENCE 227 AA; 26307 MW; 94C5347D530C55AE CRC64;
SQ
Query Match 1.2%; Score 7; DB 1; Length 227;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 304 SPPVNPD 310
Db 187 SPPVNPD 193
RESULT 37
TRY3_HUMAN
ID TRY3_HUMAN STANDARD; PRT; 247 AA.
AC P15951;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE TRYPSIN III PRECURSOR (EC 3.4.21.4).
GN PRSS3 OR TRY3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreas;
RX MEDLINE=90221895; PubMed=2326201;
RA Tani T., Kawashima I., Mita K., Takiguchi Y.;
RT "Nucleotide sequence of the human pancreatic trypsinogen III cDNA.";
RL Nucleic Acids Res. 18:1631-1631(1990).
CC -1- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE: ARG-, LYS-.
CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X15505; CAA33527.1; -
DR PIR; S12764; S12764.
DR HSSP; P00761; 1EPT.
DR MEROPS; S01.174; -
DR InterPro; IPR001254; -
DR InterPro; IPR001314; -
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
```

DR PROSITE; PS00135; TRYPSIN_SER; FALSE_NEG.
 KW Hydrolase; Serine protease; Digestion; Pancreas; Zymogen; Signal;
 KW Multigene family.
 FT SIGNAL 1 15
 FT PROPEP 16 23 ACTIVATION PEPTIDE.
 FT CHAIN 24 247
 FT ACT_SITE 63 63 TRYPsin III.
 FT ACT_SITE 107 107 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 200 200 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT DISULFID 30 160 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT DISULFID 48 64 BY SIMILARITY.
 FT DISULFID 139 206 BY SIMILARITY.
 FT DISULFID 171 185 BY SIMILARITY.
 FT SITE 194 194 REQUIRED FOR SPECIFICITY (BY SIMILARITY).
 SQ SEQUENCE 247 AA; 26776 MW; 697DE163F1CEE0D6 CRC64;

Query Match 1.2%; Score 7; DB 1; Length 247;
 Best Local Similarity 100.0%; Pred. No. 29;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 520 ASCPGKI 526
 DB 173 ASCPGKI 179
 ID DLTE_BACSU STANDARD; PRT; 252 AA.
 AC P39577;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE DLTE PROTEIN.
 GN DLTE OR IPA-1R.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 OX NCBI_TaxID=1423;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=95020537; PubMed=7934828;
 RA Glaser P., Kunst F., Arnaud M., Coudart M.P., Gonzales W.,
 RA Hullo M.F., Ionescu M., Lubochinsky B., Marcelino L., Mosser I.,
 RA Presecan E., Santana M., Schneider E., Schweizer J., Vertes A.,
 RA Rapoport G., Danchin A.;
 FT "Bacillus subtilis genome project: cloning and sequencing of the 97
 RT kb region from 325 degrees to 333 degrees.";
 RL Mol. Microbiol. 10:371-384(1993).
 RN [2]
 RP CHARACTERIZATION.
 RX MEDLINE=95318144; PubMed=7797557;
 RA Perego M., Glaser P., Minutello A., Strauch M.A., Leopold K.,
 RA Fischer W.;
 RT "Incorporation of D-alanine into lipoteichoic acid and wall teichoic
 acid in Bacillus subtilis. Identification of genes and regulation.";
 RL J. Biol. Chem. 270:15598-15606(1995).
 CC -1- PATHWAY: BIOSYNTHESIS OF D-ALANYL-LIPOTEICHOIC ACID.
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (PROBABLE).
 CC -1- SIMILARITY: TO 3-OXOACYL-[ACYL-CARRIER PROTEIN] REDUCTASES.
 CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
 CC (SDR) FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X73124; CAA51557.1; -

DR EMBL; Z99123; CAB15880.1; -.
 DR PIR; S39656; S39656.
 DR Subtilist; BG10547; dlTE.
 DR InterPro; IPR002198; -.
 DR Pfam; PF00106; adh_short; 1.
 DR PROSITE; PS00061; ADH_SHORT; 1.
 KW Oxidoreductase.
 FT NP_BIND 9 33 NAD OR NADP (BY SIMILARITY).
 FT ACT_SITE 152 152 BY SIMILARITY.
 SQ SEQUENCE 252 AA; 28268 MW; DDAF1BE457D5D62A CRC64;

Query Match 1.2%; Score 7; DB 1; Length 252;
 Best Local Similarity 100.0%; Pred. No. 29;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 210 KRLLLEIG 216
 DB 23 KRLLLEIG 29
 ID NODJ_RHILT STANDARD; PRT; 262 AA.
 AC P24144;
 DT 01-MAR-1992 (Rel. 21, Created)
 DT 01-MAR-1992 (Rel. 21, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE NODULATION PROTEIN J.
 GN NODJ.
 OS Rhizobium leguminosarum (biovar trifolii).
 OG Plasmid sym.
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Rhizobiaceae; Rhizobium.
 OX NCBI_TaxID=386;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ANU 843;
 RX MEDLINE=90251164; PubMed=2338917;
 RA Surin B.P., Watson J., Hamilton W.D.O., Economou A., Downie J.A.;
 RT "Molecular characterization of the nodulation gene, nodT, from two
 RT biovars of Rhizobium leguminosarum.";
 RL Mol. Microbiol. 4:245-252(1990).
 CC -1- FUNCTION: FORMS, WITH NODI, A MEMBRANE TRANSPORT COMPLEX INVOLVED
 CC IN THE NODULATION PROCESS. IT PROBABLY EXPORT A MODIFIED BETA-1,4-
 CC LINKED N-ACETYLGLUCOSAMINE OLIGOSACCHARIDE.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE
 CC (POTENTIAL).
 CC -1- SIMILARITY: BELONGS TO THE ABC-2 SUBFAMILY OF INTEGRAL MEMBRANE
 CC PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X51411; CAA35772.1; -.
 DR PIR; S08617; S08617.
 DR InterPro; IPR000412; -.
 DR PRINTS; PR00164; ABC2TRNSPORT.
 DR PROSITE; PS00890; ABC2_MEMBRANE; 1.
 KW Plasmid; Nodulation; Transport; Transmembrane; Inner membrane.
 FT TRANSMEM 35 55 POTENTIAL.
 FT TRANSMEM 60 80 POTENTIAL.
 FT TRANSMEM 125 145 POTENTIAL.
 FT TRANSMEM 148 168 POTENTIAL.
 FT TRANSMEM 177 197 POTENTIAL.
 FT TRANSMEM 236 256 POTENTIAL.
 SQ SEQUENCE 262 AA; 28033 MW; F579E9EC3AE0A052 CRC64;

Query Match 1.2%; Score 7; DB 1; Length 262;
 Best Local Similarity 100.0%; Pred. No. 30;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 93 TKAALAG 99
 |||||
 DB 123 TKAALAG 129

RESULT 40
 ID GCTB_ACIFE STANDARD; PRT; 265 AA.
 AC Q59112;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE GLUTACONATE COA-TRANSFERASE SUBUNIT B (EC 2.8.3.12) (GCT SMALL
 DE SUBUNIT).
 GN GCTB.

OS Acidaminococcus fermentans.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Sporomusa subbranch;
 OC Acidaminococcus.
 OX NCBI_TaxID=905;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-18 AND 43-65.
 RC STRAIN=ATCC 25085;
 RX MEDLINE=95045599; PubMed=7957258;
 RA Mack M., Bendrat K., Zelder O., Eckel E., Linder D., Buckel W.;
 RT "Location of the two genes encoding glutaconate coenzyme
 RT A-transferase at the beginning of the hydroxyglutarate operon in
 RT Acidaminococcus fermentans.";
 RL Eur. J. Biochem. 226:41-51(1994).
 RN [2]

RP X-RAY CRYSTALLOGRAPHY (2.55 ANGSTROMS).
 RC STRAIN=ATCC 25085;
 RX MEDLINE=97238937; PubMed=9083111;
 RA Jacob U., Mack M., Clausen T., Huber R., Buckel W., Messerschmidt A.;
 RT "Glutaconate Coa-transferase from Acidaminococcus fermentans: the
 RT crystal structure reveals homology with other Coa-transferases.";
 RL Structure 5:415-426(1997).
 CC -1- FUNCTION: CATALYZES THE TRANSFER OF THE COA MOIETY FROM ACETYL COA
 CC TO (R)-2-HYDROXYGLUTARATE AND RELATED COMPOUNDS LIKE GLUTACONATE.
 CC -1- CATALYTIC ACTIVITY: ACETYL-COA + (E)-GLUTACONATE = ACETATE +
 CC GLUTACONYL-1-COA.
 CC -1- PATHWAY: GLUTAMATE FERMENTATION.
 CC -1- SUBUNIT: HETEROOCTAMER OF FOUR A AND FOUR B SUBUNITS.
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
 CC -1- SIMILARITY: BELONGS TO THE GLUTACONATE COA-TRANSFERASE SUBUNIT B
 CC FAMILY.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 or send an email to license@isb-sib.ch).

EMBL; X81440; CAA57200.1;
 DR PDB; 1POI; 18-MAR-98.
 DR Transferase; 3D-structure.
 FT INIT_MET 0
 FT ACT_SITE 53 53
 SQ SEQUENCE 265 AA; 29035 MW; 1E7FF61B42162FB4 CRC64;

Query Match 1.2%; Score 7; DB 1; Length 265;
 Best Local Similarity 100.0%; Pred. No. 31;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 239 AVTIKAK 245
 |||||

DB 13 AVTIKAK 19

RESULT 41
 ID YN25_MYCTU STANDARD; PRT; 282 AA.
 AC P71887;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE HYPOTHETICAL 30.0 KDA PROTEIN RV2325C.

GN RV2325C OR MTCY3G12.09
 OS Mycobacterium tuberculosis.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1773;
 RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=H37RV.

RX MEDLINE=98295987; PubMed=9634230;

RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eiglmeyer K., Gas S., Barry C.E. III, Tekala F.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares S., Squires R., Sulston J.E.,
 RA Taylor K., Whitehead S., Barrell B.G.;

RT "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence.";

RL Nature 393:537-544(1998).

CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 or send an email to license@isb-sib.ch).

EMBL; Z79702; CAB02065.1;
 DR TubercuList; RV2325C;
 KW Hypothetical protein; Transmembrane.
 FT TRANSMEM 18 38 POTENTIAL.
 FT TRANSMEM 40 60 POTENTIAL.
 FT TRANSMEM 87 107 POTENTIAL.
 FT TRANSMEM 119 139 POTENTIAL.
 FT TRANSMEM 164 184 POTENTIAL.
 FT TRANSMEM 260 280 POTENTIAL.
 SQ SEQUENCE 282 AA; 29955 MW; 689CDB75AA589D4E CRC64;

Query Match 1.2%; Score 7; DB 1; Length 282;
 Best Local Similarity 100.0%; Pred. No. 32;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 95 AALAGGT 101
 |||||
 DB 98 AALAGGT 104

RESULT 42
 ID YMX8_CAEL STANDARD; PRT; 283 AA.
 AC P34516;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 01-JUN-1994 (Rel. 29, Last annotation update)

DE HYPOTHETICAL 32.6 KDA PROTEIN K06H7.8 IN CHROMOSOME III.
 GN K06H7.8.
 OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;

```

CC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RC SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE=94150719; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,
RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
RA Johnston L., Jones M., Kershaw J., Kirsten J., Laister N.,
RA LaReille P., Lighning J., Lloyd C., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
RA Sims M., Smaldon N., Smith A., Smith M., Sonhammer E., Staden R.,
RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
RA Waterson R., Watson A., Weinstock L., Wilkinson-Sproat J.,
RA Wohldman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L15314; AAA28091.1; -.
DR PIR; S44848; S44848.
DR HSSP; Q06486; ICK1.
DR WormPep; K06H7.8; CE00259.
KW Hypothetical protein.
SQ SEQUENCE 283 AA; 32597 MW; 84492C4D99984296 CRC64;
-----
Query Match 1.2%; Score 7; DB 1; Length 283;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 368 KCVASGK 374
DB 34 KCVASGK 40
-----
RESULT 43
SM30_STRPU STANDARD; PRT; 290 AA.
AC P28163;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE 30 KDA SPICULE MATRIX PROTEIN PRECURSOR.
GN SM30.
OS Strongylocentrotus purpuratus (Purple sea urchin).
OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC Echinoidea; Euechinoidea; Echinacea; Echinoida; Strongylocentrotidae;
OC Strongylocentrotus.
OX NCBI_TaxID=7668;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RX MEDLINE=92008838; PubMed=1717322;
RA George N.C., Killian C.E., Wilt F.H.;
RT "Characterization and expression of a gene encoding a 30.6-kDa
RT Strongylocentrotus purpuratus spicule matrix protein.";
RL Dev. Biol. 147:334-342(1991).
CC -----
CC -!- FUNCTION: MATRIX PROTEIN OF THE SEA URCHIN EMBRYO SPICULE. THE
CC CERTAIN ORIENTATIONS AND INHIBIT GROWTH IN OTHERS.
CC -!- TISSUE SPECIFICITY: ACCUMULATES EXCLUSIVELY IN MINERALIZED
CC TISSUES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```

```

CC -!- DEVELOPMENTAL STAGE: IS INCREASED AT MIDDLE TO LATE MESENCHYME
CC BLASTULA STAGE, LEVEL REMAINS HIGH THROUGH THE 3-DAY PLOTEUS
CC STAGE.
CC -----
CC -!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M63840; AAA30070.1; -.
DR InterPro; IPR001304; -.
DR Pfam; PF00059; lectin_c.1.
DR PROSITE; PS00615; C-TYPE LECTIN_1; FALSE_NEG.
DR PROSITE; PS50041; C-TYPE LECTIN_2; 1. Multigene family.
KW Matrix protein; Signal; Glycoprotein; Multigene family.
FT SIGNAL 1 15 POTENTIAL.
FT CHAIN 16 290 30 KDA SPICULE MATRIX PROTEIN.
FT DOMAIN 92 162 C-TYPE LECTIN.
FT DOMAIN 271 281 ARG-RICH (BASIC).
FT DOMAIN 215 219 POLY-ALA.
FT CARBOHYD 102 102 N-LINKED (GLCNAC...)(POTENTIAL).
SQ SEQUENCE 290 AA; 31806 MW; E1E3E27E40796DF2 CRC64;
-----
Query Match 1.2%; Score 7; DB 1; Length 290;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 306 PVNPDPPT 312
DB 39 PVNPDPPT 45
-----
RESULT 44
SM3A_STRPU STANDARD; PRT; 290 AA.
AC Q26646;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE 30 KDA SPICULE MATRIX PROTEIN-ALPHA PRECURSOR.
GN SM30-ALPHA.
OS Strongylocentrotus purpuratus (Purple sea urchin).
OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC Echinoidea; Euechinoidea; Echinacea; Echinoida; Strongylocentrotidae;
OC Strongylocentrotus.
OX NCBI_TaxID=7668;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Spem;
RX MEDLINE=94327638; PubMed=80511158;
RA Akasaka K., Frudakis T.N., Killian C.E., George N.C., Yamasu K.,
RA Khaner O., Wilt F.H.;
RT "Genomic organization of a gene encoding the spicule matrix protein
RT SM30 in the sea urchin Strongylocentrotus purpuratus.";
RL J. Biol. Chem. 269:20592-20598(1994).
CC -----
CC -!- FUNCTION: MATRIX PROTEIN OF THE SEA URCHIN EMBRYO SPICULE. THE
CC CERTAIN ORIENTATIONS AND INHIBIT GROWTH IN OTHERS (BY SIMILARITY).
CC -!- TISSUE SPECIFICITY: ACCUMULATES EXCLUSIVELY IN MINERALIZED
CC TISSUES (BY SIMILARITY).
CC -----
CC -!- DEVELOPMENTAL STAGE: EXPRESSED AT THE TIME OF SPICULE FORMATION IN
CC THE EMBRYO.
CC -----
CC -!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; U05962; AAB60620.1; -
DR EMBL; U05961; AAB60620.1; JOINED.
DR InterPro; IPR001304; -
DR Pfam; PF00059; lectin.C.1.
DR PROSITE; PS00615; C-TYPE-LECTIN_1; FALSE_NEG.
DR PROSITE; PS00041; C-TYPE-LECTIN_2; 1.
KW Matrix protein; Signal; Glycoprotein; Multigene family.
FT SIGNAL 1 20 POTENTIAL.
FT CHAIN 21 290 30 KDA SPICULE MATRIX PROTEIN-ALPHA.
FT DOMAIN 92 162 C-TYPE LECTIN.
FT DOMAIN 24 27 POLY-GLY.
FT DOMAIN 215 220 POLY-ALA.
FT DOMAIN 271 281 ARG-RICH (BASIC).
FT CARBOHYD 102 102 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 290 AA; 31686 MW; E89B5A5618E724D1 CRC64;

Db |||||
8 ELEALVK 14

Search completed: July 30, 2001, 11:44:44
Job time: 13f sec

Query Match 1.2%; Score 7; DB 1; Length 290;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 306 PVNPDPT 312
Db 39 PVNPDPT 45
|||||

RESULT 45
GLS1_BACSU STANDARD; PRT; 309 AA.
AC 007637;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE PROBABLE GLUTAMINASE YLAM (EC 3.5.1.2).
GN YLAM.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID:1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA Purnelle B., Presecan E., Glaser P., Richou A., Danchin A.,
RA Goffeau A.;
RT "Bacillus subtilis chromosomal region downstream nprE.";
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: L-GLUTAMINE + H(2)O = L-GLUTAMATE + NH(3).
CC -1- SIMILARITY: BELONGS TO THE GLUTAMINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z97025; CAB09718.1; -
DR EMBL; Z99111; CAB13356.1; -
DR Subtilisin; BG13350; YLAM.
KW Hydrolase.
SQ SEQUENCE 309 AA; 34012 MW; ELDFFFLA0723CIA5 CRC64;

Query Match 1.2%; Score 7; DB 1; Length 309;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 151 ELEALVK 157

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2001, 11:42:06 ; Search time 21.49 Seconds
(without alignments)
3521.567 Million cell updates/sec

Title: US-09-367-496-8
Perfect score: 572
Sequence: 1 MSFOGKKSIPRITSRLIR.....RTAQKIMAPGGRSNITSL 572

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 425026 seqs, 132305027 residues

Word size : 0
Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SPTREMBL16:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phase:*
10: sp_plant:*
11: sp_rodent:*
12: sp_unclassified:*
13: sp_vertebrate:*
14: sp_virus:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	25	4.4	574	13 Q9DDZ6	Q9ddz6 gallus gall
2	14	2.4	519	11 Q9EQF5	Q9eqf5 mus musculus
3	14	2.4	589	5 Q9V3N7	Q9v3n7 drosophila
4	10	1.7	531	10 Q9FNP3	Q9fnp3 arabisidopsis
5	9	1.6	332	5 Q9VND9	Q9vnd9 drosophila
6	9	1.6	1030	5 Q9N445	Q9n445 caenorhabdi
7	8	1.4	319	2 Q9F3A7	Q9f3a7 streptomyce
8	8	1.4	376	10 Q9SBF7	Q9sbf7 arabisidopsis
9	8	1.4	487	2 Q52642	Q52642 pseudomonas
10	8	1.4	542	3 Q9P903	Q9p903 saccharomyc
11	8	1.4	563	11 Q9JMG8	Q9jmg8 rattus norv
12	8	1.4	564	4 Q9NRY9	Q9nry9 homo sapien
13	8	1.4	564	4 Q9NQC4	Q9nqc4 homo sapien
14	8	1.4	564	11 Q9JHU0	Q9jhu0 rattus norv
15	8	1.4	564	11 Q9EQF6	Q9eqf6 mus musculus
16	8	1.4	575	5 Q9NAP6	Q9nap6 caenorhabdi
17	8	1.4	581	2 Q9RJC2	Q9rjc2 streptomyce
18	8	1.4	648	10 Q9M269	Q9m269 arabisidopsis
19	8	1.4	657	13 Q9DER8	Q9der8 fugu rubrip

20	8	1.4	787	13 Q9DEC8	Q9dec8 oncorhynchu
21	7	1.2	76	10 Q9LRF0	Q9lrf0 arabisidopsis
22	7	1.2	89	11 Q03317	Q03317 mus musculus
23	7	1.2	94	1 O28403	O28403 archaeoglob
24	7	1.2	112	2 Q9EZQ4	Q9ezq4 azoarcus sp
25	7	1.2	115	5 Q17259	Q17259 brachionus
26	7	1.2	115	14 Q9INE2	Q9ine2 human immun
27	7	1.2	119	5 Q9U4W3	Q9u4w3 plasmodium
28	7	1.2	123	2 Q9KB96	Q9kb96 bacillus ha
29	7	1.2	126	8 Q9TAJ8	Q9taj8 cafeteriar
30	7	1.2	127	2 Q9PMK0	Q9pmk0 campylobact
31	7	1.2	128	2 Q9LBT4	Q9lbt4 borrelia sp
32	7	1.2	129	2 Q9LBT3	Q9lbt3 borrelia sp
33	7	1.2	130	2 Q54544	Q54544 borrelia bu
34	7	1.2	130	2 Q54614	Q54614 borrelia bu
35	7	1.2	130	2 Q54501	Q54501 borrelia bu
36	7	1.2	130	2 Q54502	Q54502 borrelia bu
37	7	1.2	130	2 Q9LBT6	Q9lbt6 borrelia sp
38	7	1.2	135	2 Q9RVY5	Q9rvy5 deinococcus
39	7	1.2	143	2 Q9X5U7	Q9x5u7 borrelia bu
40	7	1.2	143	2 Q9RG11	Q9rg11 borrelia af
41	7	1.2	146	10 Q42474	Q42474 oryza sativ
42	7	1.2	147	14 Q98185	Q98185 molluscum c
43	7	1.2	151	2 Q31321	Q31321 borrelia bu
44	7	1.2	154	2 Q34675	Q34675 borrelia bu
45	7	1.2	155	2 Q31320	Q31320 borrelia bu

ALIGNMENTS

RESULT 1

Q9DDZ6 PRELIMINARY; PRT; 574 AA.
AC Q9DDZ6;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DE 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE COLLAPSED RESPONSE MEDIATOR PROTEIN 3.
GN CRMP3.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=RETINA;
RX MEDLINE=20545548; PubMed=10956643;
RA Fukada M., Watakabe I., Yuasa-Kawada J., Kawachi H., Kuroiwa A.,
RA Matsuda Y., Noda M.;
RT "Molecular Characterization of CRMP5, a Novel Member of the Collapsin
RT Response Mediator Protein Family."
RL J. Biol. Chem. 275:37957-37965(2000).
DR EMBL; AF249294; AAG37997.1; -. 96E17AC0E96CA5F4 CRC64;
SQ SEQUENCE 574 AA; 62070 MW; 96E17AC0E96CA5F4 CRC64;

Query Match 4.4%; Score 25; DB 13; Length 574;

Best Local Similarity 100.0%; Pred. No. 1.5e-16; Mismatches 0; Indels 0; Gaps 0;

QY 466 GRFVPRKTFDFVYVKRIKARNLAE 490

|||||
Db 466 GRFVPRKTFDFVYVKRIKARNLAE 490

RESULT 2

Q9EQF5 PRELIMINARY; PRT; 519 AA.
ID Q9EQF5;
AC Q9EQF5;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)

RA	Reinhart K., Remington K., Saunders R.D.C., Scheeler F., Shen H., Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T., RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E., RA Svirkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X., Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J., Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A., Ye J., Yeh R.F., Zaveril J.S., Zhan M., Zhang G., Zhao Q., Zheng L., Zhong H.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O., RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.; RT "The genome sequence of Drosophila melanogaster." RL Science 287:2185-2195(2000).
RL	[2]
RP	SEQUENCE FROM N.A.
RA	Tsang G., Brokstein P., Frise E., Harvey D., Evans-Holm M., RA Lewis S.E., Suh C., Rubin G.M.;
RT	*Full Length Drosophila melanogaster cDNA sequence."
RL	Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
RP	[3]
RP	SEQUENCE FROM N.A.
RA	Wang L.H., Strittmatter S.M.;
RT	"Drosophila CRMP is expressed in the developing nervous system.;"
RL	Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
DR	EMBL; AF003602; AAF52002.1; -
DR	EMBL; AF160900; AAD46840.1; -
DR	EMBL; AF077837; AAD52007.1; -
DR	FlyBase; FBgn0023023; CRMP.
DR	InterPro; IPR002195; -
DR	Pfam; PF00744; Dihydroorotase; 1.
SQ	SEQUENCE 589 AA; 64430 MW; 959AA4B657CF3891 CRC64;
Query Match	2.4%; Score 14; DB 5; Length 589;
Best Local Similarity	100.0%; Pred. No. 2.7e-05;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps	
OY	380 FVAVTSTNAKIFN 393
Dd	384 FVAVTSTNAKIFN 397
RESULT 4	
Q9FMP3	
ID Q9FMP3	PRELIMINARY; PRT; 531 AA.
AC Q9FMP3;	
DT 01-MAR-2001 (TREMBLrel. 16, Created)	
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)	
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)	
DE DIHROPYRIMIDINASE.	
OS Arabidopsis thaliana (Mouse-ear cress).	
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;	
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;	
OC Brassicales; Brassicaceae; Arabidopsi.	
OX NCBI_TaxID=3702;	
XP [1]	
RP SEQUENCE FROM N.A.	
RC STRAIN-COLOMBIA;	
RX MEDLINE=98162728; PubMed=9501997;	
RA Takamura Y., Sato S., Kaneko T., Kotani H., Asamizu E., Miyajima N., RA Nabata S.;	
RT "Structural analysis of Arabidopsis thaliana chromosome 5. III.	
RT Sequence features of the regions of 1,191,918 bp covered by seventeen	
RT Physically assigned Pl clones."	
RL DNA Res. 4:401-414(1997).	
DR EMBL; AB007727; BAB10038.1; -	
SQ SEQUENCE 531 AA; 57991 MW; F6DE523D73027789 CRC64;	
Query Match	1.7%; Score 10; DB 10; Length 531;
Best Local Similarity	100.0%; Pred. No. 0.3;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps	
OY	214 ELGTGPEGH 223

Db 242 ELGITGPEGH 251

```

RESULT 5
Q9VND9 PRELIMINARY; PRT; 332 AA.
AC Q9VND9;
DT 01-MAR-2001 (TREMblrel. 13, Created)
DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)
DE 01-MAR-2001 (TREMblrel. 16, Last annotation update)
DE CRMP PROTEIN.
GN CRMP OR BCDNA:HL02693 OR CG1411.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Eukaryota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celiker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew K.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Kocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Fertiera S., Fleischmann W.,
RA Foaier C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jaiswal M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasako P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclab J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL; AE003602; AAF52003.2; -.
DR FlyBase; FBgn0023023; CRMP.
DR InterPro; IPR002195; -.
DR Pfam; PF00744; Dihydroorotase; 1.
SQ SEQUENCE 332 AA; 36453 MW; 97BF4D2CFF3F601 CRC64;

```

```

Query Match 1.6%; Score 9; DB 5; Length 332;
Best Local Similarity 100.0%; Pred. No. 2;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 228 PEVEAEAV 236
DB 235 PEVEAEAV 243

```

```

RESULT 6
Q9N445 PRELIMINARY; PRT; 1030 AA.
AC Q9N445;
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
DT 01-MAR-2001 (TREMblrel. 16, Last annotation update)
DE Y37ELLAL.C PROTEIN.
GN Y37ELLAL.C.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Waterston R.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC024757; AAF59446.1; -.
DR InterPro; IPR000994; -.
DR InterPro; IPR001714; -.
DR Pfam; PF00557; Peptidase_M24; 1.
DR PRINTS; PR00599; MAPEPTIDASE.
SQ SEQUENCE 1030 AA; 115902 MW; 57058054CE8B60ED CRC64;

Query Match 1.6%; Score 9; DB 5; Length 1030;
Best Local Similarity 100.0%; Pred. No. 5.9;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 39 DGLIKQIGE 47
DB 36 DGLIKQIGE 44

RESULT 7
Q9F3A7 PRELIMINARY; PRT; 319 AA.
AC Q9F3A7;
DT 01-MAR-2001 (TREMblrel. 16, Created)
DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)
DT 01-MAR-2001 (TREMblrel. 16, Last annotation update)
DE PUTATIVE REGULATORY PROTEIN.
GN SC5F1.22C.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Saunders D.C., Harris D.;
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Cerdeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
RA Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RX MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kieser H.M., Denapante D., Eichner A., Cullum J.,

```

RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
the 8 Mb streptomycetes coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
DR EMBL; AL450165; CAC16448.1;
SQ SEQUENCE 319 AA; 32573 MW; 3A35803084F8D8AC CRC64;

Query Match 1.4%; Score 8; DB 2; Length 319;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 559 MAPPGGRS 566
|||||||
Db 171 MAPPGGRS 178

RESULT 8
Q9SBF7 PRELIMINARY; PRT; 376 AA.
AC Q9SBF7;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE PUTATIVE TRANSCRIPTION FACTOR.
GN MYB68.
OS Arabidopsis thaliana (Mouse-ear cross).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RX MEDLINE=99056848; PubMed=9839469;
RA Kranz H.D., Denekamp M., Greco R., Jin H.-L., Leyva A., Weissner R.,
RA Petroni K., Urzainqui A., Bevan M., Martin C., Smeekens S.,
RA Tonelli C., Paz-Ares J., Weisshaar B.;
RT "Towards functional characterisation of the members of the R2R3-MYB
RT gene family from Arabidopsis thaliana.";
RL Plant J. 16:263-276(1998).
DR EMBL; AF062901; AAC83623.1;
DR HSSP; P01103; 1POM.
DR InterPro; IPR001005;
DR Pfam; PF00249; myb_DNA-binding; 2.
DR PROSITE; PS00037; MYB.1; UNKNOWN_1.
DR PROSITE; PS00334; MYB.2; UNKNOWN_1.
DR PROSITE; PS00090; MYB.3; 2.
DR SMART; SM00395; SANT; 1.
SQ SEQUENCE 376 AA; 42084 MW; 342FF295D5D5DED22 CRC64;

Query Match 1.4%; Score 8; DB 10; Length 376;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 95 AALAGGTT 102
|||||||
Db 337 AALAGGTT 344

RESULT 9
O52642 PRELIMINARY; PRT; 487 AA.
AC O52642;
DT 01-JUN-1998 (Tremblrel. 06, Created)
DT 01-JUN-1998 (Tremblrel. 06, Last sequence update)
DT 01-MAY-2000 (Tremblrel. 13, Last annotation update)
DE NITROBENZALDEHYDE DEHYDROGENASE NTNC.
GN NTNC.
OS Pseudomonas sp.
OC Bacteria; Proteobacteria.
OX NCBI_TaxID=306;

RA SEQUENCE FROM N.A.
RC STRAIN=TW3;
RX MEDLINE=98215169; PubMed=9555884;
RA James K.D., Williams P.A.;
RT "ntn genes determining the early steps in the divergent catabolism of
4-nitrotoluene and toluene in Pseudomonas sp. strain TW3.";
RL J. Bacteriol. 180:2043-2049(1998).
DR EMBL; AF043544; AAC38357.1;
DR HSSP; P56533; 1BPW.
DR InterPro; IPR002086;
DR Pfam; PF00171; aldedh; 1.
DR PROSITE; PS00070; ALDEHYDE_DEHYDR_GLU; UNKNOWN_1.
DR PROSITE; PS00687; ALDEHYDE_DEHYDR_GLU; UNKNOWN_1.
SQ SEQUENCE 487 AA; 51806 MW; 6E2FDAF4BC4B8497 CRC64;

Query Match 1.4%; Score 8; DB 2; Length 487;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 65 VLPGGVDV 72
|||||||
Db 208 VLPGGVDV 215

RESULT 10
Q9P903 PRELIMINARY; PRT; 542 AA.
AC Q9P903;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE DIHYDROPYRIMIDINASE (EC 3.5.2.2).
GN PYD2.
OS Saccharomyces kluyveri (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4934;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NRRL Y-12651;
RX MEDLINE=20123992; PubMed=10656811;
RA Gojkovic Z., Jahnke K., Schnackerz K.D., Piskur J.;
RT "PYD2 encodes 5,6-dihydropyrimidine amidohydrolase, which participates
RT in a novel fungal catabolic pathway.";
RL J. Mol. Biol. 295:1073-1087(2000).
DR EMBL; AF158967; AAF69237.1;
DR InterPro; IPR002195;
DR Pfam; PF00744; Dihydroorotase; 1.
KW Hydrolase.
SQ SEQUENCE 542 AA; 60240 MW; E8758452CBE86E98 CRC64;

Query Match 1.4%; Score 8; DB 3; Length 542;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 405 SDADLVIV 412
|||||||
Db 446 SDADLVIV 453

RESULT 11
Q9JMG8 PRELIMINARY; PRT; 563 AA.
AC Q9JMG8;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE DIHYDROPYRIMIDINASE-RELATED PROTEIN.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Yanagi S., Inatome R., Yamamura H.;
 RT "Dihydropyrimidinase-related protein.";
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AB029432; BAA89475.1; -
 DR InterPro: IPR002195; -
 DR Pfam: PF00744; Dihydroorotase; 1.
 SQ SEQUENCE 563 AA; 61395 MW; B36767BF1440004D CRC64;

Query Match 1.4%; Score 8; DB 11; Length 563;
 Best Local Similarity 100.0%; Pred. No. 35;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 215 LGITGPEG 222
 |||||
 DB 207 LGITGPEG 214

RESULT 12

O9NRY9 PRELIMINARY; PRT; 564 AA.
 AC O9NRY9;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
 DE COLLAPSPIN RESPONSE MEDIATOR PROTEIN-5.
 GN CRMP5.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Yu Z., Kryzer T.J., Griesmann G.E., Lennon V.A.;
 RT "Collapsin response mediator protein-5 (CRMP5), a novel autoantigen
 associated with paraneoplastic neurological disorders.";
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF157634; AAF80348.1; -
 DR InterPro: IPR002195; -
 DR Pfam: PF00744; Dihydroorotase; 1.
 SQ SEQUENCE 564 AA; 61459 MW; 949947A983735340 CRC64;

Query Match 1.4%; Score 8; DB 4; Length 564;
 Best Local Similarity 100.0%; Pred. No. 36;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 215 LGITGPEG 222
 |||||
 DB 208 LGITGPEG 215

RESULT 13

O9NQC4 PRELIMINARY; PRT; 564 AA.
 AC O9NQC4;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
 DE HYPOTHETICAL 61.4 KDA PROTEIN.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-SPINAL CORD;
 RA Horiuchi M., Betz H.;
 RT "Human homologue of Ulip6.";

RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AJ251275; CAB95124.1; -
 DR InterPro: IPR002195; -
 DR Pfam: PF00744; Dihydroorotase; 1.
 SQ SEQUENCE 564 AA; 61394 MW; F370DD35B0B49935 CRC64;

Query Match 1.4%; Score 8; DB 4; Length 564;
 Best Local Similarity 100.0%; Pred. No. 36;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 215 LGITGPEG 222
 |||||
 DB 208 LGITGPEG 215

RESULT 14

O9JHU0 PRELIMINARY; PRT; 564 AA.
 AC O9JHU0;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
 DE ULIP-LIKE PROTEIN.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Horiuchi M., Betz H.;
 RT "Ulip and dihydropyrimidinase like protein.";
 RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AJ131436; CAB95193.1; -
 DR InterPro: IPR002195; -
 DR Pfam: PF00744; Dihydroorotase; 1.
 SQ SEQUENCE 564 AA; 61540 MW; E078B9002F54975E CRC64;

Query Match 1.4%; Score 8; DB 11; Length 564;
 Best Local Similarity 100.0%; Pred. No. 36;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 215 LGITGPEG 222
 |||||
 DB 208 LGITGPEG 215

RESULT 15

O9EQF6 PRELIMINARY; PRT; 564 AA.
 AC O9EQF6;
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
 DE COLLAPSPIN RESPONSE MEDIATOR PROTEIN 5.
 GN CRMP5.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-SWISS WEBSTER/NIH; TISSUE-WHOLE EMBRYO;
 EX MEDLINE-20545548; PubMed-10956643;
 RA Fukada M., Wataabe I., Yuasa-Kawada J., Kawachi H., Kuroiwa A.,
 RA Matsuda Y., Noda M.;
 RT "Molecular Characterization of CRMP5, a Novel Member of the Collapsin
 Response Mediator Protein Family.";
 RL J. Biol. Chem. 275:37957-37965(2000).
 DR EMBL: AF249295; AAG37998.1; -
 SQ SEQUENCE 564 AA; 61516 MW; CA93790FC8F9CD98 CRC64;

```

Query Match          1.4%; Score 8; DB 11; Length 564;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 215 LGITGPEG 222
Db 208 LGITGPEG 215
|||||||

RESULT 16
Q9NAP6 PRELIMINARY; PRT; 575 AA.
AC Q9NAP6;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE K09E4.4 PROTEIN.
GN K09E4.4.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Smye R.;
RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology.";
RL Science 282:2012-2018(1998).
DR EMBL; 283234; CAB70170.1; -.
SQ SEQUENCE 575 AA; 66849 MW; FE7216DB39F4D93C CRC64;

Query Match          1.4%; Score 8; DB 5; Length 575;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 211 RLLELGIT 218
Db 98 RLLELGIT 105
|||||||

RESULT 17
Q9RJ2 PRELIMINARY; PRT; 581 AA.
AC Q9RJ2;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE HYPOTHETICAL 63.8 KDA PROTEIN.
GN SCF91.02C.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Oliver K., Harris D.;
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Thomson N.R., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.

RC STRAIN=A3(2);
RX MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kieser H.M., Denapaita D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
DR EMBL; AL132973; CAB61159.1; -.
DR InterPro; IPR001440; -.
DR Pfam; PF00515; TPR; 3.
KW Hypothetical protein.
SQ SEQUENCE 581 AA; 63847 MW; 65F1A06AEBE3D53C CRC64;

Query Match          1.4%; Score 8; DB 2; Length 581;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 232 EAEAVYRA 239
Db 494 EAEAVYRA 501
|||||||

RESULT 18
Q9M269 PRELIMINARY; PRT; 648 AA.
ID Q9M269;
AC Q9M269;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE SERINE/THREONINE-PROTEIN KINASE-LIKE PROTEIN.
GN F21F14.130.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Choisne N., Robert C., Brottier P., Wincker P., Cattolico L.,
RA Altiguenave F., Saurin W., Weissenbach J., Mewes H.W., Lemcke K.,
RA Mayer K.F.X., Quefrier F., Salanoubat M.;
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project.
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: TO THE SER/THR FAMILY OF PROTEIN KINASES.
DR EMBL; AL138642; CAB71903.1; -.
DR InterPro; IPR000719; -.
DR InterPro; IPR001245; -.
DR InterPro; IPR002290; -.
DR Pfam; PF00069; pkinase; 1.
DR PRINTS; PR00109; TYRKINASE.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; UNKNOWN_1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR SMART; SM00220; S_TK; 1.
KW ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase;
KW Tyrosine-protein kinase.
SQ SEQUENCE 648 AA; 72131 MW; 20808421BF6A3C55 CRC64;

Query Match          1.4%; Score 8; DB 10; Length 648;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 307 VNPDPPTA 314
Db 515 VNPDPPTA 522
|||||||

RESULT 19

```

Tue Jul 31 13:08:34 2001

```

Q9DER8
ID Q9DER8 PRELIMINARY; PRT; 657 AA.
AC Q9DER8;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE TRANSPORT ASSOCIATED PROTEIN.
GN ABCB3B.
OS Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
OC Tetraodontidae; Takifugu.
OX NCBI_TaxID=31033;
RN [1]
RN SEQUENCE FROM N.A.
RX MEDLINE=20491935; PubMed=11035083;
RA Clark M.S., Pontarotti P., Gilles A., Kelly A., Elgar G.;
RT "Identification and characterization of a beta proteasome subunit
cluster in the Japanese pufferfish (Fugu rubripes).";
RL J. Immunol. 165:4446-4452(2000).
DR EMBL; AJ271723; CAC13121.1; -.
SQ SEQUENCE 657 AA; 72750 MW; B657D0CAAF61945E CRC64;

Query Match 1.4%; Score 8; DB 13; Length 657;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 445 VVISQGRV 452
DB 612 VVISQGRV 619

RESULT 20
Q9DEC8 PRELIMINARY; PRT; 787 AA.
AC Q9DEC8;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE COMPLEMENT FACTOR B/C2-B.
GN BFC2-B.
OS Oncorhynchus mykiss (Rainbow trout) (Salmo gairdneri).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.
OX NCBI_TaxID=8022;
RN [1]
RN SEQUENCE FROM N.A.
RX TISSUE=LIVER;
RA Nakao M., Fujiki K., Yano T., Bayne C.J.;
RT "Molecular cloning of B/C2-B, an isotype of complement factor B/C2,
from teleost.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB044939; BAB19788.1; -.
SQ SEQUENCE 787 AA; 87578 MW; 9111C72F80816B0C CRC64;

Query Match 1.4%; Score 8; DB 13; Length 787;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 537 SGFSLSGS 544
DB 136 SGFSLSGS 143

RESULT 21
Q9LFF0 PRELIMINARY; PRT; 76 AA.
AC Q9LFF0;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE HYPOTHETICAL 8.2 KDA PROTEIN.
GN FSE19.10.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RN SEQUENCE FROM N.A.
RX Sato S., Nakamura Y., Kaneko T., Kato T., Asamizu E., Kotani H.,
Tabata S., Mewes H.W., Rudd S., Lemcke K., Mayer K.F.X.;
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RN SEQUENCE FROM N.A.
RX EU Arabidopsis sequencing project;
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL391147; CAC01831.1; -.
KW Hypothetical protein.
SQ SEQUENCE 76 AA; 8169 MW; F60F2CA8BBEF137A CRC64;

Query Match 1.2%; Score 7; DB 10; Length 76;
Best Local Similarity 100.0%; Pred. No. 57;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 230 EYEAFAV 236
DB 62 EYEAFAV 68

RESULT 22
Q03317 PRELIMINARY; PRT; 89 AA.
AC Q03317;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
DE HYPOTHETICAL 9.3 KDA PROTEIN (ORF1).
GN TCEAL.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RN SEQUENCE FROM N.A.
RX MEDLINE=83153686; PubMed=3346229;
RA Kanai A., Kuzuhara T., Sekimizu K., Natori S.;
RT "Molecular cloning and characterization of cDNA for eukaryotic
transcription factor S-II.";
RL J. Biochem. 109:674-677(1991).
DR EMBL; D00925; BAA00767.1; -.
DR MGD; MGI:1196624; Tceal.
KW Hypothetical protein.
SQ SEQUENCE 89 AA; 9348 MW; 3EF3976B06857986 CRC64;

Query Match 1.2%; Score 7; DB 11; Length 89;
Best Local Similarity 100.0%; Pred. No. 66;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 449 OGRVALE 455
DB 66 OGRVALE 72

RESULT 23
Q28403 PRELIMINARY; PRT; 94 AA.
AC Q28403;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-JAN-1998 (TREMBlrel. 05, Last annotation update)

```

DT 01-AUG-1998 (TREMBLrel. 07, Last annotation update)
 DE CONSERVED HYPOTHETICAL PROTEIN.
 GN AF1876.
 OS Archaeoglobus fulgidus.
 OC Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;
 OC Archaeoglobus.
 OX NCBI_TaxID=2234;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
 RX MEDLINE=98049343; PubMed=9389475;
 RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
 RA Ketchum K.A., Dodson R.J., Winn M., Hickey E.K., Peterson J.D.,
 RA Richardson D.L., Kervage A.R., Graham D.E., Kyprides N.C.,
 RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
 RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
 RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
 RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
 RA Cotton M.D., Spriggs T., Artach P., Kaine B.P., Sykes S.M.,
 RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
 RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
 RA Venter J.C.;
 RT "The complete genome sequence of the hyperthermophilic, sulphate-
 RT reducing archaeon Archaeoglobus fulgidus.";
 RL Nature 390:364-370(1997).
 DR EMBL; AE000973; AAB89376.1; -.
 DR TIGR; AF1876; -.
 KW Hypothetical protein.
 SQ SEQUENCE 94 AA; 10657 MW; FDD6FB00B313918E CRC64;

Query Match 1.2%; Score 7; DB 1; Length 94;
 Best Local Similarity 100.0%; Pred. No. 69;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 155 LVKEGV 161
 Db 85 LVKEGV 91
 |||||

RESULT 24
 Q9EZQ4
 ID Q9EZQ4 PRELIMINARY; PRT; 112 AA.
 AC Q9EZQ4;
 DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
 DE P2-LIKE SIGNAL TRANSMITTER PROTEIN GLNY.
 GN GLNY.
 OS Azoarcus sp. BH72.
 OC Bacteria; Proteobacteria; beta subdivision; Rhodocyclus group;
 OC Azoarcus.
 OX NCBI_TaxID=62928;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BH72;
 RX MEDLINE=20521930; PubMed=11069654;
 RA Martin D.E., Hurek T., Reinhold-Hurek B.;
 RT "Occurrence of three PII-like signal transmitter proteins in the
 RT diazotrophic proteobacterium Azoarcus sp. BH72.";
 RL Mol. Microbiol. 38:276-288(2000).
 DR EMBL; AF281015; AAG40186.1; -.
 SQ SEQUENCE 112 AA; 12150 MW; C18A32878387C16A CRC64;

Query Match 1.2%; Score 7; DB 2; Length 112;
 Best Local Similarity 100.0%; Pred. No. 82;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 456 DGKMFVT 462
 Db 88 DGKMFVT 94
 |||||

RESULT 25
 Q17259
 ID Q17259 PRELIMINARY; PRT; 115 AA.
 AC Q17259;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
 DE CHAPERONIN (FRAGMENT).
 GN HSP60.
 OS Brachionus plicatilis.
 OC Eukaryota; Metazoa; Rotifera; Monogononta; Ploumida; Brachionidae;
 OC Brachionus.
 OX NCBI_TaxID=10195;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Cochran B.J., Mattley Y.M., Snell T.W.;
 RL Environ. Toxicol. Chem. 13:1221-1229(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Cochran B.J.;
 RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
 DR EMBL; U20804; AAA64304.1; -.
 DR InterPro; IPR002423; -.
 DR Pfam; PF00118; cpa60_TCP1; 1.
 FT NON_TER 1
 FT NON_TER 115
 SQ SEQUENCE 115 AA; 12697 MW; 5A24CD3F43B87CDE CRC64;

Query Match 1.2%; Score 7; DB 5; Length 115;
 Best Local Similarity 100.0%; Pred. No. 84;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 229 EEVEAEA 235
 Db 84 EEVEAEA 90
 |||||

RESULT 26
 Q9INE2
 ID Q9INE2 PRELIMINARY; PRT; 115 AA.
 AC Q9INE2;
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
 DE GAG PROTEIN (FRAGMENT).
 GN GAG.
 OS Human immunodeficiency virus type 1.
 OC Viruses; Retroviruses; Retroviridae; Lentivirus.
 OX NCBI_TaxID=11676;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=DRCM1.
 RX MEDLINE=99260287; PubMed=10331444;
 RA Mokill J.L., Wade C.M., Burns S.M., Cutting W.A., Bopopi J.M.,
 RA Green S.D., Peutherer J.F., Simmonds P.;
 RT "Genetic heterogeneity of HIV type 1 subtypes in Kimpese, rural
 RT Democratic Republic of Congo.";
 RL AIDS Res. Hum. Retroviruses 15:655-664(1999).
 CC -1- MISCELLANEOUS: GAG GENE PROTEIN P17 FOUND N-TERMINAL TO GAG P24
 CC (CORE NUCLEOCAPSID PROTEIN) IN THE GAG POLYPROTEIN (BY
 CC SIMILARITY).
 CC -1- SIMILARITY: TO RETROVIRAL MATRIX PROTEIN (GAG GENE PROTEIN P17).
 CC EMBL; AF144807; AAF69033.1; -.
 DR InterPro; IPR000071; -.
 DR Pfam; PF00540; gag_p17; 1.
 DR PRINTS; PR00234; HIVMATRIX.
 KW AIDS: Core protein; Polyprotein.
 FT NON_TER 1
 FT NON_TER 115
 SQ SEQUENCE 115 AA; 12712 MW; 83A6FAF50A9CC648 CRC64;

Query Match 1.2%; Score 7; DB 14; Length 115;
Best Local Similarity 100.0%; Pred. No. 84;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 203 DIVEEQ 209

DB 68 DIVEEQ 74

RESULT 27

Q9U4W3 PRELIMINARY; PRT; 119 AA.
AC Q9U4W3
DT 01-MAY-2000 (TREMELREL. 13, Created)
DT 01-MAY-2000 (TREMELREL. 13, Last sequence update)
DT 01-MAR-2001 (TREMELREL. 16, Last annotation update)
DE DBL ALPHA PROTEIN (FRAGMENT).
CN VAR.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3D7;
RX MEDLINE=20078856; PubMed=10613695;
RA Taylor H.M., Kyes S.A., Harris D., Kriek N., Newbold C.I.;
RT "A study of var gene transcription in vitro using universal var gene
primers."
RL Mol. Biochem. Parasitol. 105:13-23(2000).
DR EMBL; AF133889; AAD33650.1; -.
FT NON_TER 1
FT NON_TER 119
SQ SEQUENCE 119 AA; 13775 MW; A16DAFA9C31615N4 CRC64;

Query Match 1.2%; Score 7; DB 5; Length 119;
Best Local Similarity 100.0%; Pred. No. 87;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 355 TNGIEER 361

DB 41 TNGIEER 47

RESULT 28

Q9KB96 PRELIMINARY; PRT; 123 AA.
AC Q9KB96
DT 01-OCT-2000 (TREMELREL. 15, Created)
DT 01-OCT-2000 (TREMELREL. 15, Last sequence update)
DT 01-OCT-2000 (TREMELREL. 15, Last annotation update)
DE BH2032 PROTEIN.
GN BH2032.
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=86665;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C-125 / JCM 9153;
RA Takami H., Nakasone K., Takaki Y.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP001514; BAB05751.1; -.
SQ SEQUENCE 123 AA; 14253 MW; D232028D8E8148EC CRC64;

Query Match 1.2%; Score 7; DB 2; Length 123;
Best Local Similarity 100.0%; Pred. No. 89;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 261 ADAIAQA 267

||||||

Db 75 ADAIAQA 81

RESULT 29

Q9TAJ8 PRELIMINARY; PRT; 126 AA.
AC Q9TAJ8
DT 01-MAY-2000 (TREMELREL. 13, Created)
DT 01-MAY-2000 (TREMELREL. 13, Last sequence update)
DT 01-JUN-2000 (TREMELREL. 14, Last annotation update)
DE RIBOSOMAL PROTEIN S12.
GN RPS12.
OS Cafeteria roenbergensis.
OC Mitochondrion.
OC Eukaryota; stramenopiles; Bicosocida; Cafeteria.
OX NCBI_TaxID=33653;
RN [1]
RP SEQUENCE FROM N.A.
RA Burger G.;
RT "The mitochondrial genome of Cafeteria roenbergensis."
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF193903; AAF05788.1; -.
DR InterPro; IPR000230; -.
DR Pfam; PF00164; Ribosomal_S12; 1.
DR PRINTS; PRJ1034; RIBOSOMAL_S12.
DR PROSITE; PS00055; RIBOSOMAL_S12; UNKNOWN_1.
DR Ribosomal protein; Mitochondrion.
KW RIBOSOMAL PROTEIN; Mitochondrion.
SQ SEQUENCE 126 AA; 14019 MW; 4C43D31F0D7C7889 CRC64;

Query Match 1.2%; Score 7; DB 8; Length 126;
Best Local Similarity 100.0%; Pred. No. 91;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 LIRGGRI 24

DB 82 LIRGGRI 88

RESULT 30

Q9PMK0 PRELIMINARY; PRT; 127 AA.
AC Q9PMK0
DT 01-OCT-2000 (TREMELREL. 15, Created)
DT 01-OCT-2000 (TREMELREL. 15, Last sequence update)
DT 01-MAR-2001 (TREMELREL. 16, Last annotation update)
DE HYPOTHETICAL PROTEIN CJ1460.
GN CJ1460.
OS Campylobacter jejuni.
OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;
OC Campylobacter.
OX NCBI_TaxID=197;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCTC 11168;
RX MEDLINE=20150912; PubMed=10688204;
RA Parkhill J., Wren B.W., Mungall K., Ketley J.M., Churcher C.,
RA Basham D., Chillingworth T., Davies R.M., Feltwell T., Holtroyd S.,
RA Jagels K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W.,
RA Quail M.A., Rajandream M.A., Rutherford K.M., Van Vliet A.H.M.,
RA Whitehead S., Barrell B.G.;
RT "The genome sequence of the food-borne pathogen Campylobacter jejuni
reveals hypervariable sequences."
RL Nature 403:665-668(2000).
DR EMBL; ALJ39078; CAB73883.1; -.
KW Hypothetical protein.
SQ SEQUENCE 127 AA; 14666 MW; 3AF4F6B43DF472F4 CRC64;

Query Match 1.2%; Score 7; DB 2; Length 127;
Best Local Similarity 100.0%; Pred. No. 92;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 208 EQRLLLE 214
 Db 85 EQRLLLE 91

RESULT 31
 Q9LBT4
 ID Q9LBT4 PRELIMINARY; PRT; 128 AA.
 AC Q9LBT4;
 DT 01-OCT-2000 (Tremblrel. 15, Created)
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
 DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
 DE FLAGELLIN (FRAGMENT).
 GN FLA.
 OS Borrelia sp. TMI.
 OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
 OX NCBI_TaxID=113255;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-TMI;
 RA Masuzawa T., Kudeken M., Ochi A.;
 RT "Borrelia TMI flagellin gene."
 RL Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
 DR EMBL; AB037129; BAA90306.1; -;
 DR InterPro: IPR001444; -;
 DR Pfam: PF00460; flg_bb_rod; 1.
 DR PROSITE; PS00588; FLAGELLA_BB_ROD; UNKNOWN_1.
 FT NON_TER 1
 FT NON_TER 128
 SQ SEQUENCE 128 AA; 13114 MW; E31B76B94478BC72 CRC64;

Query Match 1.2%; Score 7; DB 2; Length 128;
 Best Local Similarity 100.0%; Pred. No. 93;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 540 SLGSGQA 546
 Db 37 SLGSGQA 43

RESULT 32
 Q9LBT3
 ID Q9LBT3 PRELIMINARY; PRT; 129 AA.
 AC Q9LBT3;
 DT 01-OCT-2000 (Tremblrel. 15, Created)
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
 DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
 DE FLAGELLIN (FRAGMENT).
 GN FLA.
 OS Borrelia sp. TAL.
 OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
 OX NCBI_TaxID=113256;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-TAL;
 RA Masuzawa T., Kudeken M., Ochi A.;
 RT "Borrelia TAL flagellin gene."
 RL Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
 DR EMBL; AB037130; BAA90307.1; -;
 DR InterPro: IPR001444; -;
 DR Pfam: PF00460; flg_bb_rod; 1.
 DR PROSITE; PS00588; FLAGELLA_BB_ROD; UNKNOWN_1.
 FT NON_TER 1
 FT NON_TER 129
 SQ SEQUENCE 129 AA; 13212 MW; 26B03A44B94478BC CRC64;

Query Match 1.2%; Score 7; DB 2; Length 129;
 Best Local Similarity 100.0%; Pred. No. 93;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 540 SLGSGQA 546

Db 37 SLGSGQA 43

RESULT 33
 O54544
 ID O54544 PRELIMINARY; PRT; 130 AA.
 AC O54544;
 DT 01-JUN-1998 (Tremblrel. 06, Created)
 DT 01-JUN-1998 (Tremblrel. 06, Last sequence update)
 DT 01-MAY-2000 (Tremblrel. 13, Last annotation update)
 DE FLAGELLIN PROTEIN (FRAGMENT).
 GN FLA.
 OS Borrelia burgdorferi (Lyme disease spirochete).
 OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
 OX NCBI_TaxID=139;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Norris D.E., Johnson B.J.B., Piesman J., Maupin G.O., Clark J.L.,
 RA Black W.C. IV;
 RL Submitted (MAR-1998) to the EMBL/GenBank/DDBJ databases.
 DR EMBL; U96235; AAC09048.1; -;
 DR EMBL; U96235; AAC09047.1; -;
 DR InterPro: IPR001444; -;
 DR Pfam: PF00460; flg_bb_rod; 1.
 DR PROSITE; PS00588; FLAGELLA_BB_ROD; UNKNOWN_1.
 FT NON_TER 1
 FT NON_TER 130
 SQ SEQUENCE 130 AA; 13723 MW; 711326BFB9C05E7 CRC64;

Query Match 1.2%; Score 7; DB 2; Length 130;
 Best Local Similarity 100.0%; Pred. No. 94;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 540 SLGSGQA 546
 Db 63 SLGSGQA 69

RESULT 34
 O54614
 ID O54614 PRELIMINARY; PRT; 130 AA.
 AC O54614;
 DT 01-JUN-1998 (Tremblrel. 06, Created)
 DT 01-JUN-1998 (Tremblrel. 06, Last sequence update)
 DT 01-MAY-2000 (Tremblrel. 13, Last annotation update)
 DE FLAGELLIN PROTEIN (FRAGMENT).
 GN FLA.
 OS Borrelia burgdorferi (Lyme disease spirochete).
 OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
 OX NCBI_TaxID=139;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Norris D.E., Johnson B.J.B., Piesman J., Maupin G.O., Clark J.L.,
 RA Black W.C. IV;
 RL Submitted (MAR-1998) to the EMBL/GenBank/DDBJ databases.
 DR EMBL; U96238; AAC09050.1; -;
 DR EMBL; U96237; AAC09049.1; -;
 DR InterPro: IPR001444; -;
 DR Pfam: PF00460; flg_bb_rod; 1.
 DR PROSITE; PS00588; FLAGELLA_BB_ROD; UNKNOWN_1.
 FT NON_TER 1
 FT NON_TER 130
 SQ SEQUENCE 130 AA; 13779 MW; EF0BAFBA395AB2 CRC64;

Query Match 1.2%; Score 7; DB 2; Length 130;
 Best Local Similarity 100.0%; Pred. No. 94;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 540 SLGSGQA 546

* Tue Jul 31 13:08:34 2001

us-09-367-496-8.oli.rspt

Db 63 SLSGSQA 69

RESULT 35

O54501 PRELIMINARY; PRT; 130 AA.

AC O54501, 1998 (Tremblrel. 06, Created)
 DT 01-JUN-1998 (Tremblrel. 06, Last sequence update)
 DT 01-JUN-1998 (Tremblrel. 13, Last annotation update)
 DE FLAGELLIN PROTEIN (FRAGMENT).

GN FLA.

OS Borrelia burgdorferi (Lyme disease spirochete).

OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.

OX NCBI_TaxID=139;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=JD-1;

RA Norris D.E., Johnson B.J.B., Piesman J., Maupin G.O., Clark J.L.,

RA Black W.C. IV.;

RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.

DR EMBL; U96234; AAC09046.1; -.

DR InterPro; IPR001444; -.

DR Pfam; PF00460; flg_bb_rod; 1.

DR PROSITE; PS00588; FLAGELLA_BB_ROD; UNKNOWN_1.

FT NON_TER 1

FT NON_TER 130 130

SQ SEQUENCE 130 AA; 13735 MW; 9238BD2FAA395095 CRC64;

Query Match 1.2%; Score 7; DB 2; Length 130;

Best Local Similarity 100.0%; Pred. No. 94;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 540 SLSGSQA 546

|||||

Db 63 SLSGSQA 69

RESULT 36

O54502 PRELIMINARY; PRT; 130 AA.

AC O54502;
 DT 01-JUN-1998 (Tremblrel. 06, Created)
 DT 01-JUN-1998 (Tremblrel. 06, Last sequence update)
 DT 01-MAY-2000 (Tremblrel. 13, Last annotation update)
 DE FLAGELLIN PROTEIN (FRAGMENT).

GN FLA.

OS Borrelia burgdorferi (Lyme disease spirochete).

OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.

OX NCBI_TaxID=139;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=CO94-27/SP10;

RA Norris D.E., Johnson B.J.B., Piesman J., Maupin G.O., Clark J.L.,

RA Black W.C. IV.;

RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.

DR EMBL; U96239; AAC09051.1; -.

DR InterPro; IPR001444; -.

DR Pfam; PF00460; flg_bb_rod; 1.

DR PROSITE; PS00588; FLAGELLA_BB_ROD; UNKNOWN_1.

FT NON_TER 1

FT NON_TER 130 130

SQ SEQUENCE 130 AA; 13751 MW; 061326BFB89C1D50 CRC64;

Query Match 1.2%; Score 7; DB 2; Length 130;

Best Local Similarity 100.0%; Pred. No. 94;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 540 SLSGSQA 546

|||||

Db 63 SLSGSQA 69

RESULT 37

O9LBT6 PRELIMINARY; PRT; 130 AA.

AC O9LBT6;
 DT 01-OCT-2000 (Tremblrel. 15, Created)
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
 DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
 DE FLAGELLIN (FRAGMENT).

GN FLA.

OS Borrelia sp. KRI.

OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.

OX NCBI_TaxID=113253;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=KRI;

RA Masuzawa T., Kudeken M., Ochi A.;

RT "Borrelia KRI flagellin gene.";

RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL; AB07127; BAA90304.1; -.

DR InterPro; IPR001444; -.

DR Pfam; PF00460; flg_bb_rod; 1.

DR PROSITE; PS00588; FLAGELLA_BB_ROD; UNKNOWN_1.

FT NON_TER 1

FT NON_TER 130 130

SQ SEQUENCE 130 AA; 13300 MW; 93D6A31B76B94478 CRC64;

Query Match 1.2%; Score 7; DB 2; Length 130;

Best Local Similarity 100.0%; Pred. No. 94;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 540 SLSGSQA 546

|||||

Db 37 SLSGSQA 43

RESULT 38

O9RVY5 PRELIMINARY; PRT; 135 AA.

AC O9RVY5;
 DT 01-MAY-2000 (Tremblrel. 13, Created)
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
 DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
 DE CONSERVED HYPOTHETICAL PROTEIN.

GN DR0886.

OS Deinococcus radiodurans.

OC Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.

OX NCBI_TaxID=1299;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=RJ;

RX MEDLINE=20036896; PubMed=10567266;

RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,

RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,

RA Mofat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,

RA Vamathevan J.J., Lam P., McDonald L., Otterback T., Zaleski C.,

RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,

RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,

RA Fraser C.M.;

RT "Genome sequence of the radioresistant bacterium Deinococcus

RL Science 286:1571-1577(1999).

DR EMBL; AE001941; AAF10458.1; -.

DR TIGR; DR0886; -.

SQ SEQUENCE 135 AA; 14523 MW; CF3546EF6A9FB68B CRC64;

Query Match 1.2%; Score 7; DB 2; Length 135;

Best Local Similarity 100.0%; Pred. No. 98;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 65 VLPGGVD 71
Db 101 VLPGGVD 107

RESULT 39
Q9X5U7
ID Q9X5U7 PRELIMINARY; PRT; 143 AA.
AC Q9X5U7
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE FLAGELLIN PROTEIN (FRAGMENT).
OS Borrelia burgdorferi group.
OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
OX NCBI_TaxID=64895;
RN [1]
RP SEQUENCE FROM N.A.
RA Leutenegger C.M., Pusterla N., Mislin C., Lutz H.;
RT "Molecular evidence of Ticks infected with Borrelia burgdorferi sensu
latu in Switzerland."
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF127531; AAD31190.1; -
DR InterPro; IPR001444; -
DR Pfam; PF00460; flg_bb_rod; 1.
DR PROSITE; PS00588; FLAGELLIN_BB_ROD; UNKNOWN_1.
FT NON_TER 1
FT NON_TER 143
FT SEQUENCE 143 AA; 14913 MW; 733C3EB0171C4C8B CRC64;

Query Match 1.2%; Score 7; DB 2; Length 143;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 540 SLSGSQ 546
Db 44 SLSGSQ 50

RESULT 40
Q9RG11
ID Q9RG11 PRELIMINARY; PRT; 143 AA.
AC Q9RG11
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE FLAGELLIN PROTEIN (FRAGMENT).
OS Borrelia afzelii
OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
OX NCBI_TaxID=29518;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=99419177; PubMed=10488215;
RA Leutenegger C.M., Pusterla N., Mislin C.N., Weber R., Lutz H.;
RT "Molecular evidence of coinfection of ticks with Borrelia burgdorferi
sensu lato and the human granulocytic ehrlichiosis agent in
Switzerland."
RL J. Clin. Microbiol. 37:3390-3391(1999).
DR EMBL; AF127532; AAD31191.1; -
DR InterPro; IPR001444; -
DR Pfam; PF00460; flg_bb_rod; 1.
FT NON_TER 1
FT NON_TER 143
FT SEQUENCE 143 AA; 14872 MW; 98A5BB44B2972B0C CRC64;

Query Match 1.2%; Score 7; DB 2; Length 143;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 540 SLSGSQ 546
Db 44 SLSGSQ 50

Db 44 SLSGSQ 50

RESULT 41
Q42474
ID Q42474 PRELIMINARY; PRT; 146 AA.
AC Q42474
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE ROOT-SPECIFIC PROTEIN (PUTATIVE LIPID TRANSFER PROTEIN).
GN RC2 OR RCC2.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartioideae; Oryzaceae;
OC Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=GULFMONT; TISSUE=ROOT;
RX MEDLINE=95195153; PubMed=7888615;
RA Xu Y., Buchholz W.G., DeRose R.T., Hall T.C.;
RT "Characterization of a rice gene family encoding root-specific
proteins."
RL Plant Mol. Biol. 27:237-248(1995).
RN [2]
RP SEQUENCE FROM N.A.
RA STRAIN=CV. NIPPONBARE;
RA Hsiao C.R., Yuan Q., Moffat K.S., Hill J.N., Jenkins C.N., Burr P.C.,
RA Bowman C.L., Zismann V., Pai G., Bowman C.L., Fujii C.Y., Vanaken S.E.,
RA Quackenbush J., Craven B., Utterback T.R., Khalak H., Feldblyum T.V.,
RA Quackenbush J., White O., Salzberg S.L., Fraser C.M.;
RT "Oryza sativa chromosome 10 BAC OSJNBa0015J15 genomic sequence."
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; L27210; AAA79836.1; -
DR EMBL; L27209; AAA65512.1; -
DR HSSP; P24337; IHYP.
DR Mendel; 9716; Oryza; 1531; 9716.
DR InterPro; IPR000528; -
DR Pfam; PF00279; LTP; 1.
DR SEQUENCE 146 AA; 14466 MW; E51E2E89F193457F CRC64;

Query Match 1.2%; Score 7; DB 10; Length 146;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 441 GAPAVVI 447
Db 32 GAPAVVI 38

RESULT 42
Q98185
ID Q98185 PRELIMINARY; PRT; 147 AA.
AC Q98185; 012881; 012600;
DT 01-FEB-1997 (TREMBlrel. 02, Created)
DT 01-FEB-1997 (TREMBlrel. 02, Last sequence update)
DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)
DE MC014R.
GN MC014R OR B-M,N,L.4.
OS Molluscum contagiosum virus subtype 1 (MCV1).
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC Molluscipoxvirus.
OX NCBI_TaxID=10280;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=96325459; PubMed=8670425;
RA Senkevich T.G., Bugert J.J., Sisler J.R., Koonin E.V., Darai G.,
RA Moss B.;
RT "Genome sequence of a human tumorigenic poxvirus: prediction of
specific host response-evasion genes.";

* Tue Jul 31 13:08:34 2001

```

RL Science 273:813-816(1996).
RN [2]
RP SEQUENCE FROM N.A.
RA Senkevich T.G., Bugert J.J., Sisler J.R., Koonin E.V., Darai G.,
RA Moss B.;
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=97093414; PubMed=8938976;
RA Martin-Gallardo A., Moratilla M., Funes J.M., Agronayor M., Nunez A.,
RA Varas A.J., Collado M., Valencia A., Lopez-Estebarez J.L.,
RA Esteban M.;
RT "Sequence analysis of a Molluscum contagiosum virus DNA region which
RT includes the gene encoding protein kinase 2 and other genes with
RT unique organization.";
RL Virus Genes 13:19-29(1996).
RN [4]
RP SEQUENCE FROM N.A.
RA Moratilla M., Agronayor M., Nunez A., Funes J.M., Varas A.J.,
RA Lopez-Estebarez J.L., Esteban M., Martin-Gallardo A.;
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RA Moratilla M., Agronayor M., Nunez A., Funes J.M., Varas A.J.,
RA Lopez-Estebarez J.L., Collado M., Pavon M., Martin-Gallardo A.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; U60315; AAC55142.1; -
DR EMBL; U68894; AAB57934.1; -
KW Hypothetical protein.
SQ SEQUENCE 147 AA; 16225 MW; AF6750D93EBA2F1B CRC64;

Query Match 1.2%; Score 7; DB 14; Length 147;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 560 APPGGRS 566
DB 120 APPGGRS 126

RESULT 43
O31321 PRELIMINARY; PRT; 151 AA.
AC O31321
DT 01-JAN-1998 (TEMBLrel. 05, Created)
DT 01-JAN-1998 (TEMBLrel. 05, Last sequence update)
DT 01-MAY-2000 (TEMBLrel. 13, Last annotation update)
DE FLAGELLIN (FRAGMENT).
GN FLA.
OS Borrelia burgdorferi (Lyme disease spirochete).
OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
OX NCBI_TaxID=139;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-UK-SCO;
RA Curtin S.M., Maggs A.D.F., Carter P.E., Pennington T.H.;
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; Y15100; CAA75375.1; -
DR InterPro; IPR001029; -
DR InterPro; IPR001444; -
DR Pfam; PF00460; flg_bb_rod; 1.
DR Pfam; PF00700; Flagellin_C; 1.
DR PROSITE; PS00588; FLAGELLA_BB_ROD; UNKNOWN_1.
FT NON_TER 151
FT SEQUENCE 151 AA; 15697 MW; 4B28C6E0A0ACC9E CRC64;

Query Match 1.2%; Score 7; DB 2; Length 151;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 540 SLSGSCA 546
DB 6 SLSGSCA 12

RESULT 44
O34675 PRELIMINARY; PRT; 154 AA.
AC O34675
DT 01-JAN-1998 (TEMBLrel. 05, Created)
DT 01-JAN-1998 (TEMBLrel. 05, Last sequence update)
DT 01-MAY-2000 (TEMBLrel. 13, Last annotation update)
DE FLAGELLIN (FRAGMENT).
GN FLA.
OS Borrelia burgdorferi (Lyme disease spirochete).
OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
OX NCBI_TaxID=139;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-IP3, AND CHIN23;
RA Curtin S.M., Maggs A.D.F., Carter P.E., Pennington T.H.;
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; Y15095; CAA75370.1; -
DR EMBL; Y15091; CAA75366.1; -
DR InterPro; IPR001029; -
DR InterPro; IPR001444; -
DR Pfam; PF00460; flg_bb_rod; 1.
DR Pfam; PF00700; Flagellin_C; 1.
DR PROSITE; PS00588; FLAGELLA_BB_ROD; UNKNOWN_1.
FT NON_TER 154
FT SEQUENCE 154 AA; 16078 MW; 69BA0FC908EC8119 CRC64;

Query Match 1.2%; Score 7; DB 2; Length 154;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 540 SLSGSCA 546
DB 7 SLSGSCA 13

RESULT 45
O31320 PRELIMINARY; PRT; 155 AA.
AC O31320
DT 01-JAN-1998 (TEMBLrel. 05, Created)
DT 01-JAN-1998 (TEMBLrel. 05, Last sequence update)
DT 01-MAY-2000 (TEMBLrel. 13, Last annotation update)
DE FLAGELLIN (FRAGMENT).
GN FLA.
OS Borrelia burgdorferi (Lyme disease spirochete).
OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
OX NCBI_TaxID=139;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-UK-OX;
RA Curtin S.M., Maggs A.D.F., Carter P.E., Pennington T.H.;
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; Y15099; CAA75374.1; -
DR InterPro; IPR001029; -
DR InterPro; IPR001444; -
DR Pfam; PF00460; flg_bb_rod; 1.
DR Pfam; PF00700; Flagellin_C; 1.
DR PROSITE; PS00588; FLAGELLA_BB_ROD; UNKNOWN_1.
FT NON_TER 155
FT SEQUENCE 155 AA; 16058 MW; 6BAB1FB4CA30BEE7 CRC64;

Query Match 1.2%; Score 7; DB 2; Length 155;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Tue Jul 31 13:08:34 2001

us-09-367-496-8.oli.rspt

Page 14

Matches	7;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
Qy	540	SLGSQA	546						
Db	7	SLGSQA	13						

Search completed: July 30, 2001, 11:44:20
Job time: 134 sec